Comparison of Annual Maximum and Peaks Over Threshold Methods with Automated Threshold Selection in Flood Frequency Analysis: A Case Study for Australia

Xiao Pan
Western Sydney University
Ataur Rahman (a.rahman@uws.edu.au)
University of Western Sydney https://orcid.org/0000-0001-7152-9006

Research Article

Keywords: Flood frequency analysis, Generalized Pareto, Generalised extreme value, threshold selection, Peaks-over-threshold

Posted Date: March 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-216479/v1

Version of Record: A version of this preprint was published at Natural Hazards on November 10th, 2021. See the published version at https://doi.org/10.1007/s11069-021-05092-y.

Comparison of annual maximum and peaks over threshold methods with automated threshold selection in flood frequency analysis: A case study for Australia

Xiao Pan [1] and Ataur Rahman [1,*]

[1]Professor, School of Engineering, Design and Built Environment, Western Sydney University, Australia

[*]Building XB, Kingswood, School of Engineering, Western Sydney University, New South Wales 2751, Australia, E-mail: a.rahman@westernsydney.edu.au

Abstract

Flood frequency analysis (FFA) enables fitting of distribution functions to observed flow data for estimation of flood quantiles. Two main approaches, Annual Maximum (AM) and peaks-over-threshold (POT) are adopted for FFA. POT approach is under-employed due to its complexity and uncertainty associated with the threshold selection and independence criteria for selecting peak flows. This study evaluates the POT and AM approaches using data from 188 gauged stations in south-east Australia. POT approach adopted in this study applies a different average numbers of events per year fitted with Generalised Pareto (GP) distribution with an automated threshold detection method. The POT model extends its parametric approach to Maximum Likelihood Estimator (MLE) and Point Moment Weighted Unbiased (PMWU) method. Generalised Extreme Value (GEV) distribution using L-moment estimator is used for AM approach. It has been found that there is a large difference in design flood estimates between the AM and POT approaches for smaller average recurrence intervals (ARI), with a median difference of 25% for 1.01 year ARI and 5% for 50 and 100 years ARIs.

Keywords: Flood frequency analysis; Generalized Pareto, Generalised extreme value; threshold selection; Peaks-over-threshold.

Introduction

Flood is one of the worst natural disasters. In flood risk assessment, a design flood is defined as a flood discharge associated with a given annual exceedance probability. Flood frequency analysis (FFA) is widely adopted in estimating design floods. In FFA, two approaches are frequently adopted: annual maximum (AM) and peaks over threshold (POT). AM approach involves the selection of maximum streamflow data from each year to form AM flood series. Bezak, Brilly and Šraj (2014) stated few shortcomings of using AM approach. During the extraction of the AM series, some relatively high flow values are excluded since AM considers only one flow value from each year. AM series also includes some smaller flow values from dry years, which are much smaller than the second or third highest flow values of some wet years. These shortcomings of the AM approach may increase the uncertainty in estimated design floods, in particular for arid regions. Many studies have suggested that the AM approach results in loss of useful flood information, which is due to extraction of a relatively smaller subset from the primary data series (Bezak, Brilly & Šraj 2014; Lang, Ouarda & Bobée 1999; Langbein 1949). Comparing with POT, AM approach is advantageous due to the straightforward exaction of the AM series.

POT series extraction aims to retain any peak value above a certain truncation or threshold level based on the need for the analysis. In other words, the overall quantities of flood data to be included in the analysis is controllable by setting a relative threshold level (Lang, Ouarda & Bobée 1999). Based on the purposes of the analysis, POT can retain infinitive data points above the selected threshold value provided that the selected data points meet the independence criteria. Many studies have suggested the advantages of using POT approach in FFA (Bezak, Brilly & Šraj 2014; Bhunya et al. 2012; Bhuyan, Jena & Bhunya 2016; Bobée et al. 1993; Durocher, Burn & Ashkar 2019; Gottschalk & Krasovskaia 2002; Herath, Prasad Basnayake & Coremans 2015; Karim, Hasan & Marvanek 2017; Madsen, Rasmussen & Rosbjerg 1997; Mostofi Zadeh et al. 2019; Pan & Rahman 2018). However, POT approach remains underemployed in Australia. The additional flexibility using POT approach is associated with additional complexity in modelling (Lang, Ouarda & Bobée 1999). POT approach, in turn, is considered a dual-domain approach, time interval and magnitude, whereas only magnitude is examined in the AM approach. The associated complexities for retaining peaks are in dual aspects (Bezak, Brilly & Šraj 2014). Considering the magnitude of the flood peaks, the lowest inter-flow value requires to be determined to exclude the recess of the flood hydrograph. Time scale-wise, duration of the flow peaks needs to be analysed to suit the independence criteria for individual flood peaks.

Besides the drawback of the threshold determination for retaining flood peaks, another notable complexity is due to the determination of the statistical threshold for convergence in fitting Generalised Pareto (GP) distribution, which is the most recommended distribution for POT approach (Bernardara et al. 2012; Durocher et al. 2018; Thompson et al. 2009). Traditional graphical method using mean residual life plot for threshold selection is subjective and is not suitable to quantify and assess the associated variance of higher quantile flood estimates. The iterative process of a different threshold for estimation is typically applied to achieve better fitting of observed flood series (Durocher et al. 2018). With the convenience of computational modelling techniques, the iterative process may be automated and systematised with selected requirements.

Many studies in Australia aimed to enhance the accuracy in both at-site and regional FFA (Haddad & Rahman 2015; Haddad et al. 2012; Haddad et al. 2011; Haddad et al. 2010; Ishak et al. 2013). Traditionally, AM series is more popular than the POT for at-site and regional FFA in Australia (Ball et al. 2016). Within the Australian context, there are only few studies focusing on the POT approach in FFA (Karim, Hasan & Marvanek 2017). Due to the uniqueness of the at-site flood characteristics and parent probability distribution being unknown, this study aims to compare the AM and POT approaches using data from a large number of gauged stations in south-east Australia. This study examines automated threshold selection for GP distribution using two different methods based on normality of difference (ND) and threshold stability (TS). Hence, the objectives of this study are three folds: (i) evaluation of the extraction of the POT series based on different threshold values or the number of events per year; (ii) evaluation of the application of automated threshold detection for GP distribution in POT approach based on ND and TS methods; and (iii) comparison of POT and AM approaches in estimating flood quantiles. The findings of this study will enhance application of the POT approach in FFA in Australia and other countries.

Methods

Peaks-over-threshold approach

The first assumption that makes the POT model viable is the confirmation of Poisson arrival or homogeneous hypothesis. Lang, Ouarda and Bobée (1999) stated that the threshold value should be sufficiently high so that this assumption is not violated. However, the selected threshold value also needs to be low enough to retain as many peak flow values as possible to exploit the advantages of the POT approach. This study applies the various values of average events per year to automating the selection of the physical threshold. Through using the method, the overall quantities of data points for the analysis can be controlled. For the site having a shorter record length, POT approach can obtain more data points than AM approach. This study applies 3, 5 and 10 in average events per year, which are denoted as POT3, POT5 and POT10, respectively. Noteworthy, the associated complexity using POT for FFA is that the approach considers two elements, magnitude and time.

Regarding the aspect of time, the second flood peak is rejected if the duration between the two peaks is smaller than the calculated value. With the consideration of the magnitude, some flood peaks are to be considered as the recession of flood only. USWRC (1976) proposed two criteria using Equation 1 to ensure the independence of the data point, where θ is duration between any two flow peaks and A is the catchment area in square miles. The recession or interflow of the floods are excluded that if minimum of discharge value between any two consecutive floods (Q_1 and Q_2) is higher than three quarter of flood peaks.

$$\theta < 5\ days + log(A)\ \&\ X_{min} > \frac{3}{4}\ Min.[Q_1, Q_2\ ] \qquad (1)$$

Program using R is applied in this study to extract the POT series by applying a physical threshold and used for GP distribution fitting with an automated statistical threshold detection based on ND and TS approach, which are denoted as POT-ND and POT-TS respectively. Procedures shown in Figure 1 are followed in this study.

As mentioned previously, the complexity associated with the POT approach is the determination of the physical threshold. Below iterative process is implemented in R code to retain flood peaks.

POT series extraction

1. Determine overall quantity of data points required for POT series.

$$\text{Size of POT series} = \lambda\ (average\ events\ per\ year) * length\ of\ record \quad (2)$$

2. Select the first maximum flood data point as the first entry.
3. Compare the second maximum flood data point to the first entry by the two selection criteria. If the second maximum flood data is successfully met with the requirements with both magnitude and time, it makes an entry to the POT series, otherwise, it is rejected. Remaining data points for the iterative process are required to meet both magnitude and duration criteria as before.
4. Repeating the above procedure until the count of the data points is equal to the requirement set by step 1.
5. Export the POT series based on λ values of 3, 5 and 10, i.e. POT3, POT5 and POT10, respectively.

Non-parametric approach

Non-parametric approach is applied to estimate observed flood quantiles based on Equation 3 using POT series. The sample quantile is defined as a weighted average of consecutive order statistics. The resulting quantile estimates are approximately median-unbiased regardless of the distribution of the sample. ARI - Q relationship is formulated for comparison against parametric approach with both POT and AM approaches.

$$Q_i(p) = (1-\gamma)x_j + \gamma x_{j+1}; where\ m = \frac{p+1}{3}\ , p_k \approx [F(x_k)] \quad (3)$$

Automating statistical threshold selection for GP distribution fitting

Traditionally, the statistical threshold for GP distribution is determined through Mean Residual Life Plot (MRLP). Coles (2001) described the process as subjective and requiring field experience to interpret the plots. However, the variance of the estimates is not quantifiable through using visual assessment. Another approach aims to estimate the parameters of multiple thresholds then fit the GP distribution. Coles (2001) discussed the theoretical basis for this approach, which through *n* equally spaced increasing threshold, the suitable estimated parameters provides critical p-value from the normality test, which provided the best-suited threshold candidate. Thompson et al. (2009) discussed the automated threshold approach for wave analysis, who applied a similar approach, POT-ND. Let $\widehat{\sigma_{u_i}}$ and $\widehat{\xi_{u_i}}$ be the estimated scale and shape factors based on the threshold value, μ_i. τ_{u_i} is calculated with *n* equally spaced increasing threshold as per Equation 4.

$$\tau_{u_i} = \widehat{\sigma_{u_i}} - \widehat{\xi_{u_i}} * \mu_i \ \ , where\ i = 1, 2, 3, \dots n \tag{4}$$

The difference between individual candidate thresholds is then formulated as per Equation 5.

$$\tau_{u_i} - \tau_{u_{i-1}}, where\ i = 2, 3, 4, \dots n \tag{5}$$

Following the below iterative process, the suitable threshold μ_i is detected for GP distribution.

1. Let μ_1 to be the lowest value of the proposed threshold and μ_{100} to be the highest value of the proposed threshold. Below constraints are applied for μ_1 and μ_{100}.

   $$If\ \lambda * \text{length of record} > 100, \text{then}\ \mu_1 = Mean\ of\ POT(\lambda)$$
   $$If\ \lambda * \text{length of record} < 100,\ then\ \mu_1 = Median\ of\ POT(\lambda)$$
   $$, where\ \mu_{100} = 98\%\ quantile\ of\ POT\ model$$

2. $\mu_1 < \mu_2 < \cdots < \mu_{100}$ to be equally spaced candidate thresholds.
3. Estimate the parameters using the individual threshold with both MLE and PWMU estimators.
4. Fit GP distribution with corresponding threshold candidates.
5. Calculate the $\tau_{u_i} - \tau_{u_{i-1}}$ and examine the normality based on mean equals to zero.
6. Iterating the processes until p-value generated from the normality test met with the critical value, p-values equal to 0.25, 0.5, 0.8 and 0.9.
7. Verify the obtained threshold value through MRLP and parameters' plots for both MLE and PWMU estimators.
8. Generating P-P, Q-Q and return level plots for both GP fitting of MLE and PWM estimators.

Coles (2001) also stated that a suitable threshold can be determined based on the interpretation of the threshold stability plot. Like the above shown iterative process, if the GP distribution is appropriate for the model, estimated shape parameter versus threshold plot should be relatively constant or linear for any larger threshold compared to selected threshold. Curceac et al (2020) proposed an empirical method to find the suitable threshold for GP fitting based on rate of change of threshold versus parameter plots. This study applies the similar approach (POT-TS) to compare the obtained results.

Following the below iterative process, the suitable threshold μ_i is detected for GP distribution.

1. Let μ_1 to be the lowest value of the proposed threshold and μ_{1000} to be the highest value of the proposed threshold. $\mu_1 = Mean\ of\ POT(\lambda),\ \mu_{100} = 98\%\ quantile\ of\ POT\ model.$
2. $\mu_1 < \mu_2 < \cdots < \mu_{1000}$ to be equally spaced candidate thresholds.
3. Estimate the parameters using the individual threshold with both MLE and PWMU estimators.
4. Fit GP distribution with corresponding threshold candidates.
5. Plot the threshold against individual shape and modified scale parameters.
6. Curve fitting using cubic spline function and extract the coordinates of the plots.
7. Set boundary using 2.5% of step i.e. 25 consecutive threshold values and calculate the rate of change of extracted coordinates from the plot.
8. Determine the lowest rate of change based on step 7.
9. Verify the obtained threshold value through MRLP and threshold versus parameter plot.
10. Generate P-P, Q-Q and return level plots for both GP fitting of MLE and PWM estimators.

Annual Maximum Series

Previous studies indicate the GEV distribution with the AM approach with L-moment estimator yields satisfactory FFA results in higher return periods (Ball et al. 2016). As a benchmark for comparison, this study applies GEV distribution to AM flood series of the selected stations.

Evaluation statistics

For verification purposes, Equations 6 to 9 are applied to all the selected stations.

$$\%\ \text{difference} = \frac{Q_{i,POT\ (\lambda)} - Q_{i,GEV-AM}}{Q_{i,GEV-AM}} * 100\% \tag{6}$$

$$RMSE_r = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_{i,GEV-AM} - Q_{i,POT(\lambda)}}{Q_{i,GEV-AM}}\right)^2} \tag{7}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Q_{i,GEV-AM} - Q_{i,POT(\lambda)})^2 \tag{8}$$

$$BIAS_r = \frac{1}{n}\sum_{i=1}^{n}\frac{Q_{i,POT\ (\lambda)} - Q_{i,GEV-AM}}{Q_{i,GEV-AM}} \tag{9}$$

Study Area and Data

This study focuses on catchments located along the south-eastern coastal region of Australia as shown in Figure 2. A total of 188 gauged stations are selected for this study. Catchment areas of the selected catchments are in the range from 3 km^2 to 1010 km^2 with an average of 343 km^2 and a median of 289 km^2. Out of 188 catchments, 76 are from NSW and 102 from Victoria. The selected catchments are unregulated and there has been no major change of land use during the stream gauging period. Records of streamflow data range 23 - 98 years, with an average of 52 years and a median of 48 years.

Results

Case study

At the beginning, the result for station 223204 is presented in greater details. This station is located in Victoria, named Deptford on Nicholson River. The streamflow record length is 56 years for this station and the sample sizes for POT3, POT5 and POT10 are 168, 280 and 560, respectively.

Figure 3 compares results obtained from AM and POT approaches. For POT-ND approach, the plot includes POT-ND-3, POT-ND-5 and POT-ND-10 models using MLE and PMWU estimators based on the p-value of 0.25, and POT10 model is used for plotting of observed flood series using Cunnane's plotting position formula as shown below.

$$ARI = \frac{N + 0.2}{m - 0.4} \tag{10}$$

Here *m* is the rank of flood peaks when they are arranged in descending order and *N* is the record length in years. It can be seen from Figure 3(a) that both AM and POT approaches provide a good fit to the observed flood data in the mid ARI ranges (5 to 20 years). At 50 and 100 years ARIs, POT-ND-5 model based on MLE provides closer fit to the observed flood series compared to other POT-ND models while the AM approach provides a poor fit. Comparing the results between various POT-ND models, at smaller ARIs, it is found that all the POT-ND models provide satisfactory fit against observed flood series. Overall, POT-ND-5 model provides the best fit to the observed flood data. In Figure 3(b), the boxplots represent the percentage difference between estimated flood quantiles by different POT-ND-GP and AM-GEV approaches.

The POT-ND series selected for station 223204 is based on the automation of threshold detection using ND method as discussed in the methodology section. To yield a satisfactory fitting using GP distribution, a critical p-value of 0.25 is applied based on normality test using an iterative process as mentioned before. Three other p-values of 0.5, 0.8 and 0.9 are also applied for verification purpose. Traditionally, MRLP is used for visual inspection of the threshold, which falls into the linear section of the mean access. However, the variance of the mean access for selection of threshold is unable to quantify based on visual inspection. Figure 4 illustrates the threshold comparison between two estimators, MLE and PMWU in MRLP using different p-values. Red vertical line in Figure 4 represents the detected threshold based on 0.25 p-value using MLE. The green vertical line indicates the detected threshold using PWMU estimator. From left to right, the first figure plots the mean access for models of POT-ND-3, POT-ND-5 and POT-ND-10, respectively. Detected thresholds for various models and estimator fit close to the linear section of MRLP. The PWMU estimator provides lower detected thresholds, in general, compared to MLE. With the advantage of the POT-ND approach, the size of the extracted time series is controllable to overcome the drawback of the AM approach. POT-ND-10 models illustrate the significant lower detected threshold compared to POT-ND-3 and POT-ND-5 models while still maintaining the validity of GP distributional assumption based on normality test with the corresponding p-value. The average of the detected threshold value in POT-ND-3 model is about 50 m^3/s then it reduces to approximately 35 m^3/s for POT-ND-5 model and 23 m^3/s for POT-ND-10 model.

Modified scale and shape parameters are also plotted along with MRLP for further verification purpose. Figure 5 plots the difference between POT-ND-3, POT-ND-5 and POT-ND-10 models. For both parametric approaches, detected thresholds are sitting on the start of the linear range. This further confirms the validity of the GP distribution fitting to the POT-ND models using detected thresholds. From left to right, POT models plot average of 3, 5 and 10 events per year. With the decrease of the detected threshold, the estimated parameters are adjusted accordingly.

Furthermore, three different types of plots are generated for each parametric estimator in POT-ND approach to verify the fitting between theoretical GP and empirical GP distribution fitting. Result for station 223204, plotted with the P-P, Q-Q and return level is shown in Figure 6(a). Comparing between MLE and PMWU's estimators, the difference is difficult to visualise using these plots. However, for both parametric approaches, the P-P and Q-Q plots provide satisfactory results between theoretical and empirical dataset. With a 95% confidence interval, both estimators show a closer match with the reference line. Comparing the results obtained based on the selected POT models it is found that flood quantiles are similar for lower ARIs. However, POT-ND-10 results in lower flood discharge on higher return periods, which is as expected. Overall, through various verification methods, the automation of threshold detection using ND method with p-value of 0.25 provides satisfactory results for station 223204.

POT-TS method is also applied for station 223204 for comparison and verification purposes. Figure 6(b), (c) and (d) represent the return level plots based on POT 3, POT 5 and POT 10 using bi-estimator with both ND and TS methods, respectively.

Figure 7 represents the impact of the selection of the critical p-values of 0.25, 0.5, 0.8 and 0.9 in the normality test. With POT-ND-10 models, the threshold value is similar if a p-value of 0.25 or 0.5 is selected. However, if a higher p-value of 0.8 or 0.9 is selected, the corresponding candidate threshold value is significantly higher. This also applies to POT-ND-3 and POT-ND-5 models, and the candidate thresholds are impacted significantly if a higher p-value is selected. Based on the threshold value from Figure 7, the variance of the flood quantile is significant, especially at higher return periods. At lower return periods, all POT-ND models fit closely with the observed flood series. With selected different p-values, smaller p-values give closer fitting to the observed flood series.

The quantiles estimated by the POT-ND-GP approach are compared with the AM-GEV approach using Equation 6, and the percent differences between the POT-ND-GP and AM-GEV approaches are summarised in Table 1. For smaller ARIs up to 5 years, the differences in estimated flood discharges among the approaches are notably higher with a mean variation up to 826% for 1.10 year ARI. At 5 years ARI, the mean difference is reduced to 1%. The mean differences are 4%, 8%, 13% and 17% for ARIs of 10, 20, 50 and 100 years.

Overall, the POT3-ND-PWMU approach provides the best match with the AM-GEV approach. Overall, at smaller ARIs, the differences between the POT-GP and AM-GEV approaches are higher; however, the differences reduce as ARI increases.

Figure 8 shows distribution of Bias (equation 9), MSE (equation 8) and RMSE (equation 7) values. In terms of Bias, the best result is found for POT-ND-3-PWMU, followed by POT-ND-5-PWMU. In terms of MSE, the best result is found for POT-ND-10-MLE, followed by

POT-ND-5-PWMU. In terms of RMSE, the best result is found for POT-ND-10-MLE, followed by POT5-PWMU.

Comparison between POT and AM approaches for the selected 188 stations

This section compares the results obtained from AM and POT approaches for the selected 188 gauged stations. For each POT model, two-parameter estimation methods are used, which are MLE and PWMU through R platform. To provide an overview of the results for the selected 188 gauged stations, we firstly tabulate the overall station count for the least percentage difference between POT and AM approaches using Equation 6. Table 2(a) breaks down the overall station count based on different POT models using ND method for selected ARI brackets while Table 2(b) is constructed based on POT-TS method.

With lower ARIs up to 2 years for both methods, POT10 model provides the least percentage difference between the POT and AM approaches where over 90% of the selected stations prefer POT10-ND model and over 75% of the selected stations prefer POT10-TS model. With the increase of the ARIs up to 10 years, POT5 model provides the least percentage difference between the POT and AM approaches with more than half of the stations prefer POT5 models for both ND and TS methods. This is in agreement with Bezak, Brilly and Šraj (2014) for lower ARIs, which applied POT5 and POT8 model to compare with AM approach. However, our study differed the result of higher ARIs flood estimates to Bezak, Brilly and Šraj (2014). With the increase of ARI, the preferred POT models are POT5 and POT3 for 20, 50 and 100 years ARIs, respectively. We have found a positive relationship between percentage difference and average events per year, i.e. reducing average events per year also decreases the percentage difference between AM and POT approaches, which agrees with Bačová-Mitková and Onderka (2010) and Robson and Reed (1999). This highlights the importance of the effects of sample size (thresholds) to the final estimates (Bačová-Mitková & Onderka 2010; Bezak, Brilly & Šraj 2014; Robson & Reed 1999).

Tables 3(a) and 3(b) show count of stations having different ranges of percentage differences between all POT-GP models and AM approaches for three ARIs (1.01, 20 and 100 years). For 100 years ARI between POT-ND and AM approach, 99 stations (out of 188 stations) have differences below 5%, 44 stations have differences in the range of 5-10%. POT-TS approach provides slightly more site count for 100 years ARI, 104 stations (out of 188 stations) have differences below 5%, 32 stations have differences in the range of 5-10%. For 1.01 year ARI, 137 out of 188 stations have differences above 25% for both POT-ND and

POT-TS approaches. Overall, this result shows that the differences between POT and AM approaches are higher for greater number of stations at smaller ARIs. On the other hand, we have found that the higher the ARIs, the lesser of the percentage differences between AM and POT approaches, which agrees with the findings by Karim, Hasan and Marvanek (2017).

Figure 9(a) and 9(b) provide the percentage difference comparison between POT-GP and AM-GEV approaches with individual ARIs based on ND and TS methods. For each tier of ARI, POT3, POT5 and POT10 with bi-estimators are included in the calculation as per below boxplot. For up to 2 years ARI, the percentage differences between the two approaches are varied significantly. As the increase of the ARI, the percentage difference is reduced and is deemed to be satisfactory. Also, we have found that in most of the sites, the obtained flood estimates in higher ARIs using POT approach is lower than the obtained results from AM approach (negative percentage difference), which oppose to the results obtained by Bezak, Brilly and Šraj (2014), but is in agreement with Önöz and Bayazit (2001) and Bačová-Mitková and Onderka (2010).

Tables 4(a) and 4(b) tabulate the statistics of the percentage differences between POT and AM approaches considering all the 188 stations. The overall median percentage difference is 65% and 33% for POT-ND and POT-TS respectively, considering all the 11 ARIs and 188 stations. For 5 to 100 years ARIs, the median differences range 3-5%. For smaller ARIs (1.01 to 2 years), the median differences are in the range of 7-259% for both ND and TS methods. Overall, the obtained results for ARIs higher than 10 years (less frequent flood), POT approach provides comparable results to AM approach.

Figures 10, 11 and 12 show the spatial distribution of the stations showing different ranges of percentage differences between the POT-ND and AM approaches for 1.01, 20 and 100 years ARIs. From these figures, it can be seen that there is little spatial coherence between the percentage difference between the POT and AM approaches, i.e. these differences are randomly distributed over the space. A further regional study may be carried out to examine the spatial differences in greater details. For 1.01 year ARI, there are a greater number of stations with difference greater than 25%, and only a few stations show the percentage difference below 10%. For 20 years ARI, the percentage differences are mainly in the range of 10 – 20%. With 100 years ARI, the percentage difference is generally smaller compared to lower ARIs and is mainly in the range of 5% - 10%.

Comparison between parametric and non-parametric approaches

The differences in quantile estimates between POT-parametric and POT-non-parametric approaches are calculated. Here the peak flow data extracted based on POT3 model is used. As can be seen in Table 5(a) for POT-ND versus POT non-parametric approach, for 100 years ARI, for 92 stations (out of 188 stations), the smallest difference between POT-parametric and POT-non-parametric methods is found for POT10 model, for 68 stations, the smallest difference is for POT5 model and for 28 stations, the smallest difference is found for POT3 model. Overall, POT10 model represents the best result where 49% to 87% stations show the best match with the POT-non-parametric method. Similarly, Table 5(b) details the comparison between POT-TS and POT non-parametric approaches. POT10 model represents the best results compared to POT non-parametric approach.

Tables 6(a) and 6(b) detail the statistics of the differences between the POT-parametric and POT-non-parametric approaches for both ND and TS methods. With smaller ARI brackets, the percentage difference is higher with a mean value of 189%. With the increase of the ARIs, the mean percentage differences are reduced, and 100 years ARI has the smallest mean difference. Figure 13(a) and 13(b) present the boxplot of percentage difference between POT-parametric and POT-non-parametric approaches. Both TS and ND methods provide comparable results in higher ARIs to non-parametric approach, which agrees with Curceac et al. (2020), except we have found a greater fluctuation with lower ARIs (less than 10 years) using both ND and TS approach. This fluctuation may be further investigated using bankfull discharge series as proposed by Karim, Hasan and Marvanek (2017), which indicated that POT approach is suitable for lower ARIs, but still it is not error free.

The difference in quantile estimates between AM-GEV and POT-non-parametric approaches are calculated for comparison. Figure 14 represents the boxplots between two approaches. POT-non-parametric approach provides satisfactory results at lower ARIs (less than 2 years ARI), and the difference between two approaches increase with the increase of ARI.

Table 7 details the statistics of the percentage differences between the AM-parametric and POT-non-parametric approaches. With smaller ARI brackets, the percentage difference using Equation 6 is higher with a mean value of 144%. With the increase of the ARIs, the mean percentage differences are reduced to 29% and 30% for 50 and 100 years ARIs, respectively.

Discussion

This study applies both POT and AM approaches to 188 gauged stations located in NSW and VIC. With the current practice in Australia regarding at-site FFA, AM approach is the preferred method and POT approach remains under-deployed. The main drawback of the AM approach is the limited data points extracted and used for FFA. This drawback can be overcome through applying POT approach. However, recent studies discussed in the introduction and method section indicates the complexity and the reason for under-deployment of POT approach. The complexity mainly is from the determination of both physical and statistical threshold using current POT methods in extracting the POT series. To sufficiently retain enough flood data while maintaining the independence is challenging. This study applies the automated routine based on ND and TS methods to ensure the extracted data points are independent through programming in R. Through applying the automated routine in this study, the overall size of the time series is controllable, which overcomes one of the drawbacks of using POT approach. This is accomplished by setting the average events per year as a user input for iteration. Another difficulty for POT approach is as mentioned before, to ensure the extracted data series are suited for GP distributional fitting. This study applies the normality test and uses the p-value as a determinant to retain flood data points for further analysis. This study also applies the examination of the threshold stability to find the minimum rate of change in threshold versus estimated parameters plots. Though at the moment, there are only limited studies focusing on automation of the statistical threshold detection in at-site FFA using POT. But this methodology is applied broadly in other areas of interest such as coastal engineering and financing. Through applying the automation of threshold detection for GP distribution from this study, FFA is performed for 4 different critical p-values accompanying with two-parameter estimation techniques associating with 3, 5 and 10 average events per year. The obtained results are also verified by recent empirical approach based on theory of threshold stability.

GEV and GP distribution are two of the recommended distributions for conducting at-site FFA. This study details the comparison between these two distributions based on at-site AM flood data from 188 gauged stations. In comparing the POT to AM approach, this study mainly focuses on the comparison between GEV distribution in the AM approach and GP distribution in POT approach. The estimated results using 6 different approaches (each POT model with bi-estimator) are compared with results from GEV distribution fitting. It is further verified through comparison using POT-TS method. Though the complexity with at-site FFA is that the true distribution family is unknown, this study explores the possibility of fitting multiple POT models with different automated threshold detection techniques for comparison with individual GEV distribution fitting. The study has found and confirmed that with the increase of the ARIs, the differences in flood quantile estimates by the AM and POT approaches are reduced for the selected sites. With ARI up to 2 years, the differences between the two approaches are notably high. The median difference is then reduced to 5% for 50 and 100 years ARIs by POT-ND method. Based on obtained results, this study recommends the use of POT10 model with both ND and TS methods for ARIs up to 1.75 years, POT5 model for ARIs of 2 to 10 years, and POT3 models for ARIs of 50 and 100 years. This study also compares the difference between parametric and non-parametric approaches for selected stations which aims to evaluate the reliability of fitting GP distribution using POT approach. Obtained results show that the POT10 and POT5 models using GP distribution gives closer match with the non-parametric approach.

Comparing between the POT and AM approaches, there are more flexibilities using POT approach but accompanying with the complexity of retaining the flood peaks with both physical and statistical thresholds. The drawback of the POT approach is overcome by controlling the size of the POT model. This study also suggested a range of boundaries for setting *n* equally spaced thresholds for normality test in at-site FFA. With p-value to be pre-set in the iterative process, the candidate threshold for GP distribution is quantifiable compared to the traditional method using MRLP. This process can be further modified to suit different selection criteria.

Conclusion

The study examines the applicability of the POT model in at-site flood frequency analysis using 188 gauged stations in NSW and Victoria states in Australia. To fit the POT series, GP distribution is used while to fit the AM series GEV distribution is used. The following conclusions are drawn from this study:

- At higher ARIs (20 to 100 years) the differences in flood quantile estimates between the POT-ND and AM approaches are much smaller than for smaller ARIs. The median difference between these two approaches is 5% for 50 and 100 years ARIs, and the difference is greater than 25% for most the stations for 1.01 year ARI.

- For smaller ARIs (1.01 to 1.75 years) POT 10 model is preferable, for medium ARIs (2 to 10 years), POT5 model is preferable and for 20 to 100 years ARIs, POT3 model is preferable.
- No spatial coherence is found between the differences in POT and AM approaches in the selected study area.
- In extracting the POT series, p-value is to be pre-set in the iterative process, making the candidate threshold for GP distribution quantifiable compared to the traditional method such as using MRLP.
- Empirical method of POT-TS is proven to be suitable based on the obtained flood quantile estimates of 188 stations in Australia.

Compliance with Ethical Standards

Authors declare that there is no conflicts of interest and the research did not involve any human participants and/or animals.

References

Bačová-Mitková, V & Onderka, M 2010, 'Analysis of extreme hydrological events on the sanube using the peak over threshold method', *Journal of Hydrology and Hydromechanics*, vol. 58, no. 2, pp. 88-101.

Ball, JE, Babister, MK, Nathan, R, Weinmann, PE, Weeks, W, Retallick, M et al. 2016, *Australian Rainfall and Runoff - A Guide to Flood Estimation*, Commonwealth of Australia, Canberra.

Bernardara, P, Mazas, F, Weiss, J, Andreewsky, M, Kergadallan, X, Benoît, M et al. 2012, 'On the two step threshold selection for over-threshold modelling', Scopus.

Bezak, N, Brilly, M & Šraj, M 2014, 'Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis', *Hydrological Sciences Journal*, vol. 59, no. 5, pp. 959-77.

Bhunya, PK, Singh, RD, Berndtsson, R & Panda, SN 2012, 'Flood analysis using generalized logistic models in partial duration series', *Journal of Hydrology*, vol. 420-421, pp. 59-71.

Bhuyan, MK, Jena, J & Bhunya, PK 2016, 'At-site flood analysis using exponential and generalized logistic models in partial duration series (PDS)', *International Journal of Engineering and Technology*, vol. 8, no. 1, pp. 501-14.

Bobée, B, Cavadias, G, Ashkar, F, Bernier, J & Rasmussen, PF 1993, 'Towards a systematic approach to comparing distributions used in flood frequency analysis', *Journal of Hydrology*, vol. 142, no. 1-4, pp. 121-36.

Coles, S 2001, *An Introduction to Statistical Modelling of Extreme Values*, Springer, London.

Curceac, S, Atkinson, PM, Milne, A, Wu, L & Harris, P 2020, 'An evaluation of automated GPD threshold selection methods for hydrological extremes across different scales', *Journal of Hydrology*, vol. 585.

Durocher, M, Burn, DH & Ashkar, F 2019, 'Comparison of Estimation Methods for a Nonstationary Index-Flood Model in Flood Frequency Analysis Using Peaks Over Threshold', *Water Resources Research*, vol. 55, no. 11, pp. 9398-416.

Durocher, M, Mostofi Zadeh, S, Burn, DH & Ashkar, F 2018, 'Comparison of automatic procedures for selecting flood peaks over threshold based on goodness-of-fit tests', *Hydrological Processes*, vol. 32, no. 18, pp. 2874-87.

Gottschalk, L & Krasovskaia, I 2002, 'L-moment estimation using annual maximum (AM) and peak over threshold (POT) series in regional analysis of flood frequencies', *Norsk Geografisk Tidsskrift*, vol. 56, no. 2, pp. 179-87.

Haddad, K & Rahman, A 2015, 'Estimation of large to extreme floods using a regionalization model', in *Landscape Dynamics, Soils and Hydrological Processes in Varied Climates*, pp. 279-92.

Haddad, K, Rahman, A, Kuczera, G & Weinmann, E 2012, 'A new regionalisation model for large flood estimation in Australia: Consideration of inter-site dependence in modelling', pp. 969-76, Scopus.

Haddad, K, Rahman, A, Weeks, W, Kuczera, G & Weinmann, PE 2011, 'Towards a new regional flood frequency analysis method for Western Australia', pp. 3788-95, Scopus.

Haddad, K, Zaman, M, Rahman, A & Shrestha, S 2010, 'Regional flood modelling: Use of Monte Carlo cross-validation for the best model selection', pp. 2831-40, Scopus.

Herath, S, Prasad Basnayake, A & Coremans, D 2015, 'Comparison of flood frequency analysis by annual maximum and peak over threshold approaches for Fitzroy River, Western Australia', pp. 682-90, Scopus.

Ishak, EH, Rahman, A, Westra, S, Sharma, A & Kuczera, G 2013, 'Evaluating the non-stationarity of australian annual maximum flood', *Journal of Hydrology*, vol. 494, pp. 134-45.

Karim, F, Hasan, M & Marvanek, S 2017, 'Evaluating annual maximum and partial duration series for estimating frequency of small magnitude floods', *Water (Switzerland)*, vol. 9, no. 7.

Lang, M, Ouarda, TBMJ & Bobée, B 1999, 'Towards operational guidelines for over-threshold modeling', *Journal of Hydrology*, vol. 225, no. 3, pp. 103-17.

Langbein, WB 1949, 'Annual floods and the partial-duration flood series', *Eos, Transactions American Geophysical Union*, vol. 30, no. 6, pp. 879-81, viewed 2020/06/12.

Madsen, H, Rasmussen, PF & Rosbjerg, D 1997, 'Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modeling', *Water Resources Research*, vol. 33, no. 4, pp. 747-57.

Mostofi Zadeh, S, Durocher, M, Burn, DH & Ashkar, F 2019, 'Pooled flood frequency analysis: a comparison based on peaks-over-threshold and annual maximum series', *Hydrological Sciences Journal*, vol. 64, no. 2, pp. 121-36.

Önöz, B & Bayazit, M 2001, 'Effect of the occurrence process of the peaks over threshold on the flood estimates', *Journal of Hydrology*, vol. 244, no. 1-2, pp. 86-96.

Pan, X & Rahman, A 2018, 'Comparison of annual maximum and peaks-over-threshold methods in flood frequency analysis', pp. 614-25, Scopus.

Robson, A & Reed, D 1999, *Statistical Procedures for Flood Frequency Estimation, Flood Estimation Handbook*, Centre for Ecology & Hydrology, Wallingford, UK.

Thompson, P, Cai, Y, Reeve, D & Stander, J 2009, 'Automated threshold selection methods for extreme wave analysis', *Coastal Engineering*, vol. 56, no. 10, pp. 1013-21.

USWRC 1976, *Guidelines for determining flood flow frequency*, US Department of the Interior, Geological Survey, Office of Water Data Coordination.

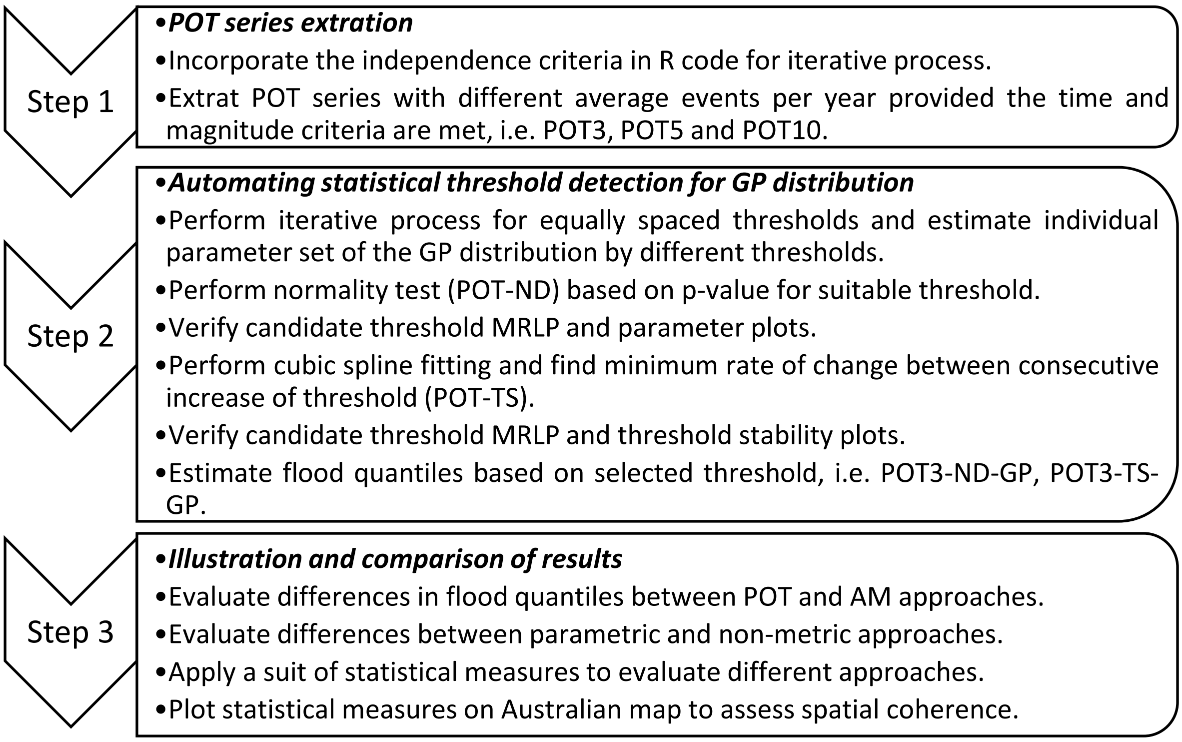


Figure 1 Illustration of adopted methodology

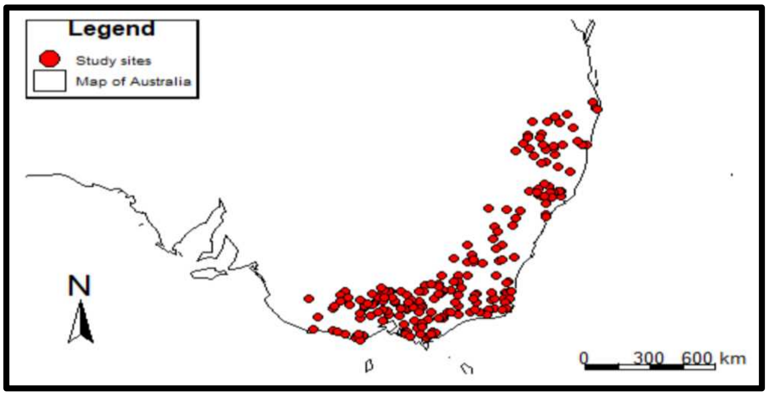


Figure 2 Distribution of selected stations

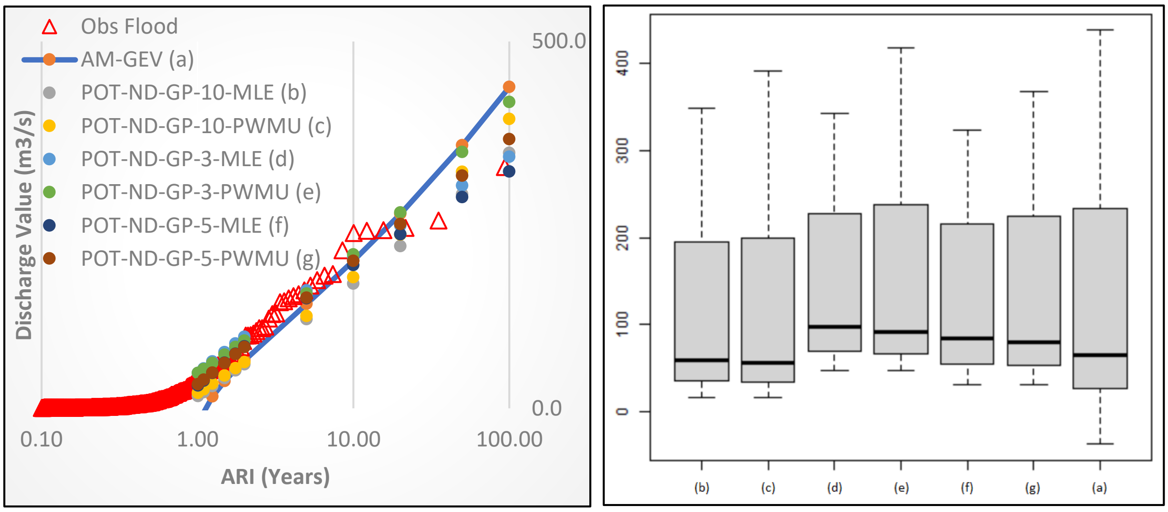


Figure 3(a) & 3(b) Comparison between POT-ND-GP (p-value=0.25) and AM-GEV approaches. Y-axis of the boxplots represents % difference estimated by equation 6.

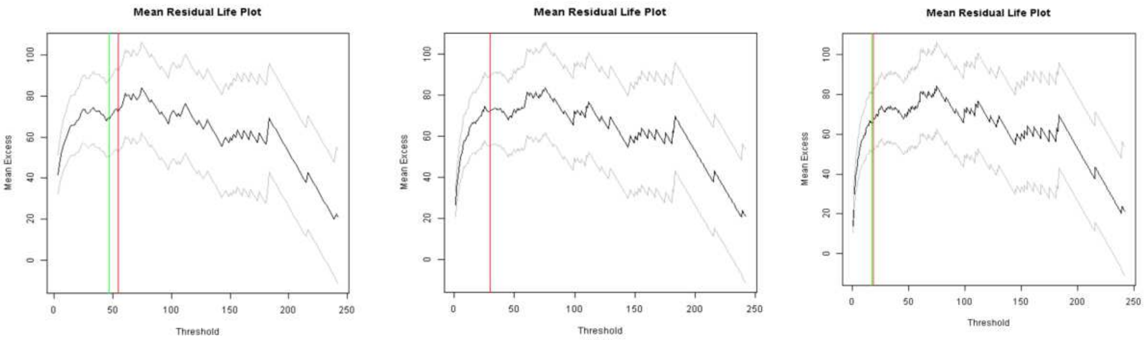


Figure 4 Mean residual life plot based on different p-values for POT-ND method

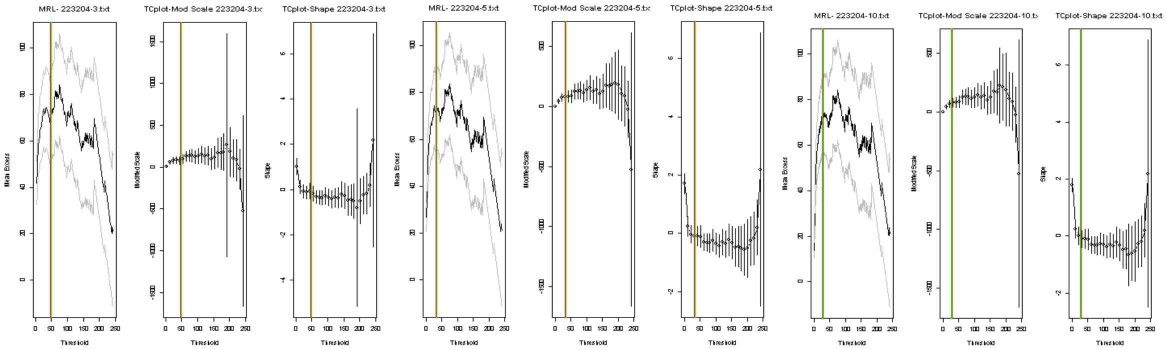


Figure 5 Estimated parameters plot with MRLP for POT-ND method

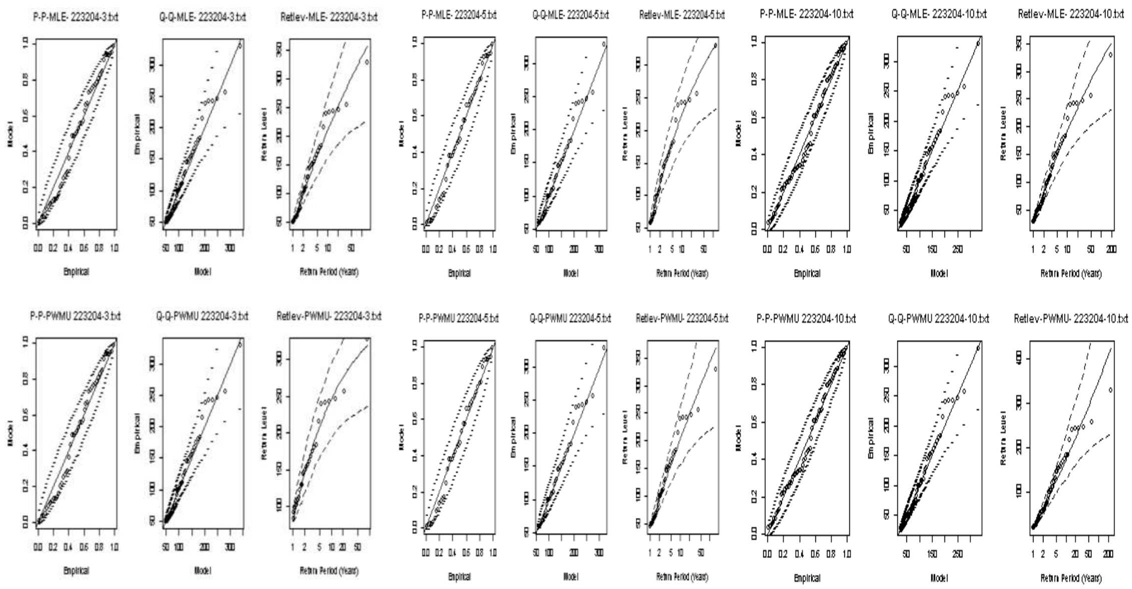


Figure 6(a) P-P, Q-Q and return level plots for different POT models using POT-ND method

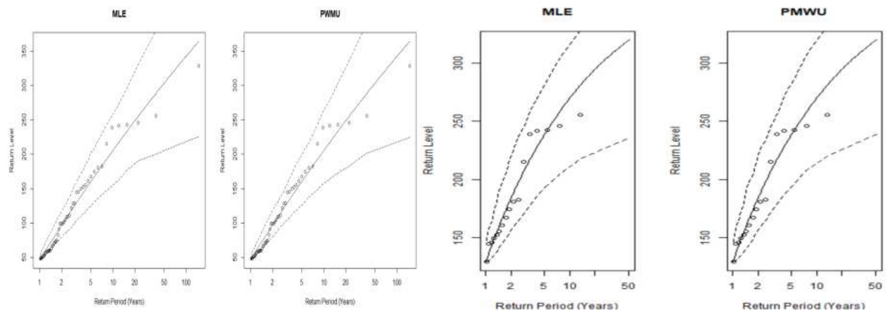


Figure 6(b) Return level comparison between POT-ND-3 and POT-TS-3. From left to right, return level plots POT-ND-3-MLE, POT-ND-3-PWMU, POT-TS-3-MLE and POT-TS-3-PWMU

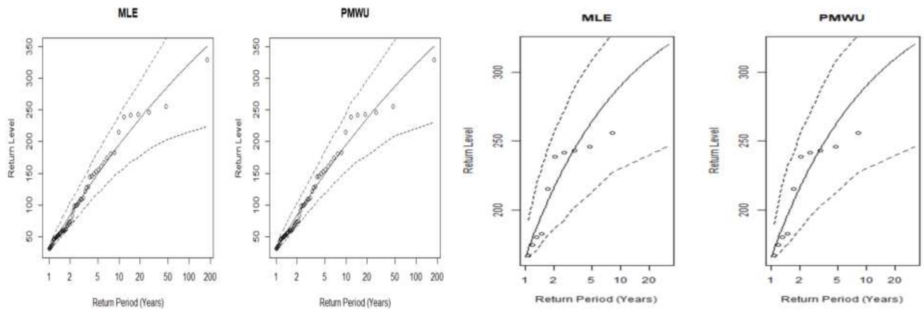


Figure 6(c) Return level comparison between POT-ND-3 and POT-TS-5. From left to right, return level plots POT-ND-5-MLE, POT-ND-5-PMWU, POT-TS-5-MLE and POT-TS-5-PWMU

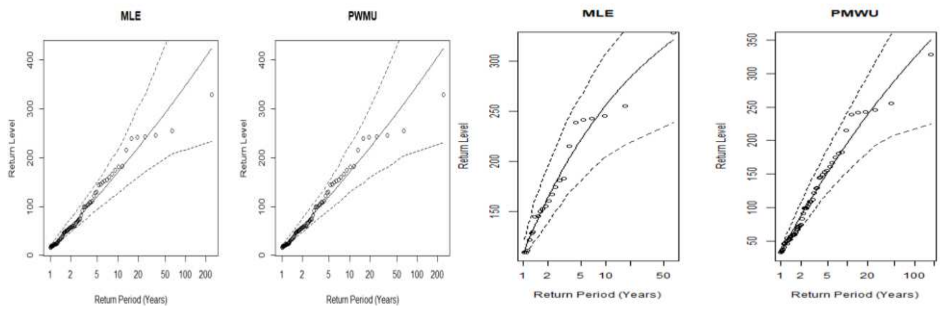


Figure 6(d) Return level comparison between POT-ND-3 and POT-TS-10. From left to right, return level plots POT-ND-10-MLE, POT-ND-10-PWMU, POT-TS-10-MLE and POT-TS-10- PWMU

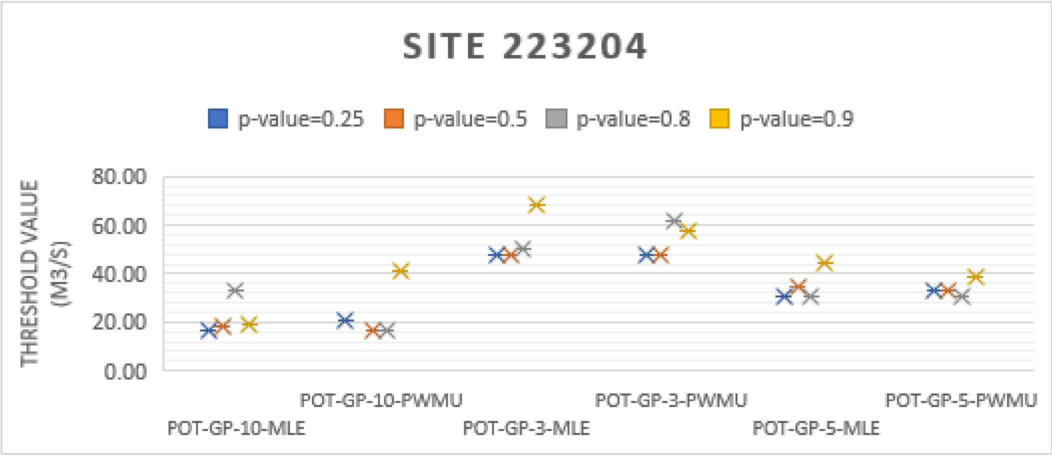
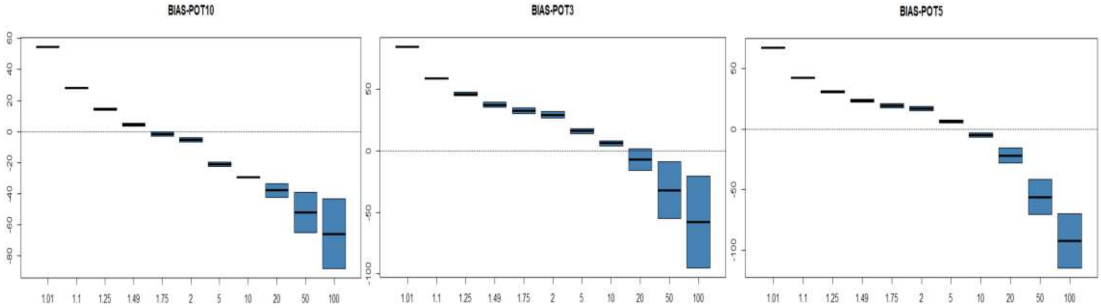
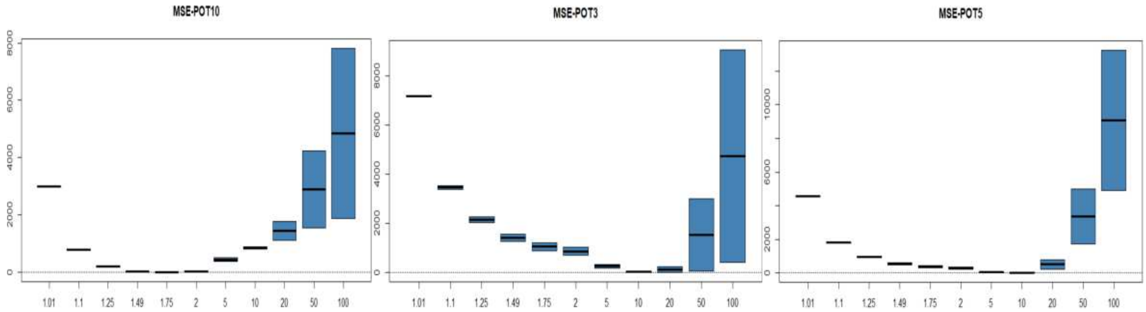


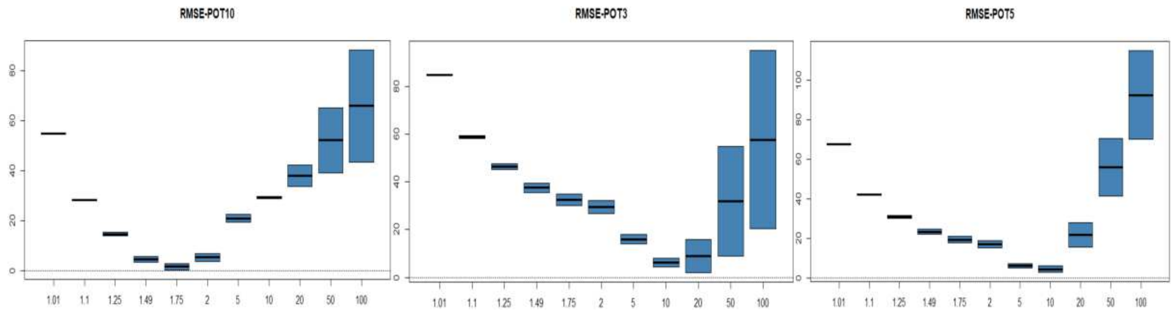
Figure 7 Threshold comparison based on varied p-value using POT-ND method



(a)



(b)



(c)

Figure 8 Verification Plot - (a) Bias Boxplot (b) MSE Boxplot (c) RMSE Box plot; from left to right, plots represent POT-ND-10, POT-ND-3 and POT-ND-5, respectively

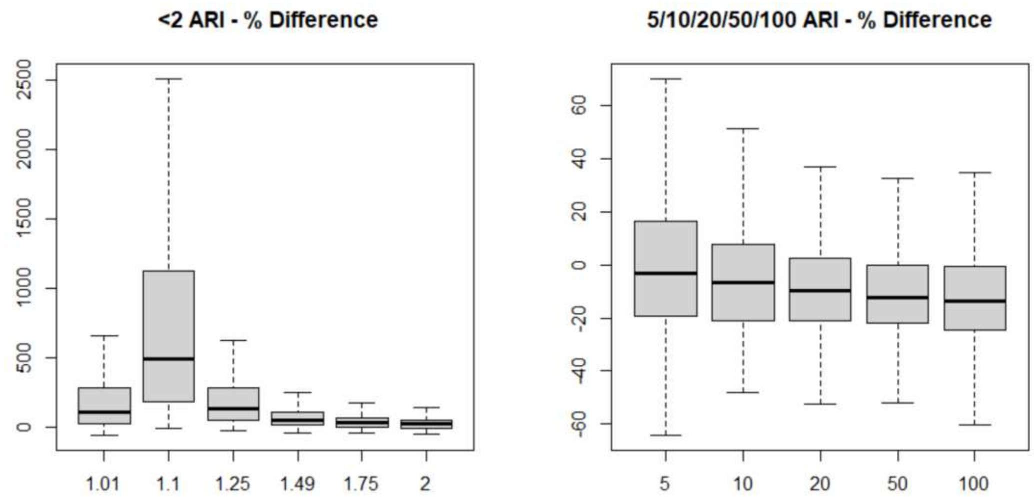


Figure 9(a) Comparison between POT-ND-GP and AM-GEV approaches – boxplots showing % difference

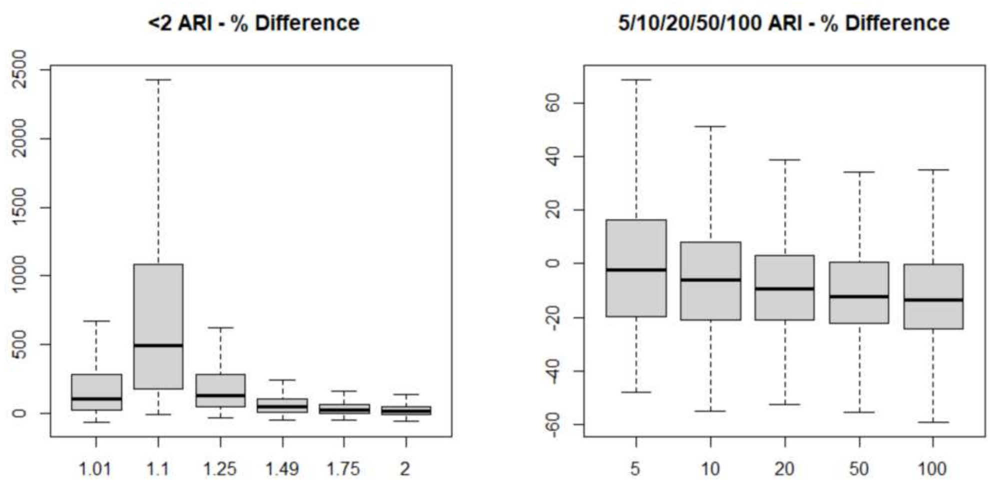


Figure 9(b) Comparison between POT-ND-GP and AM-GEV approaches – boxplots showing % difference

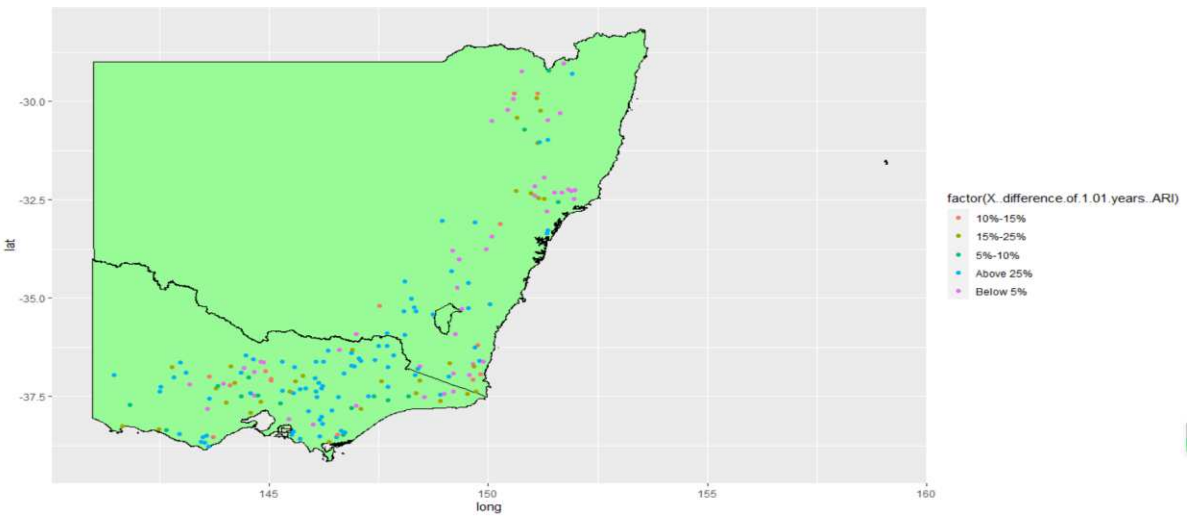


Figure 10 Geographical distribution of percentage differences between POT-ND and AM approaches (1.01 year ARI)

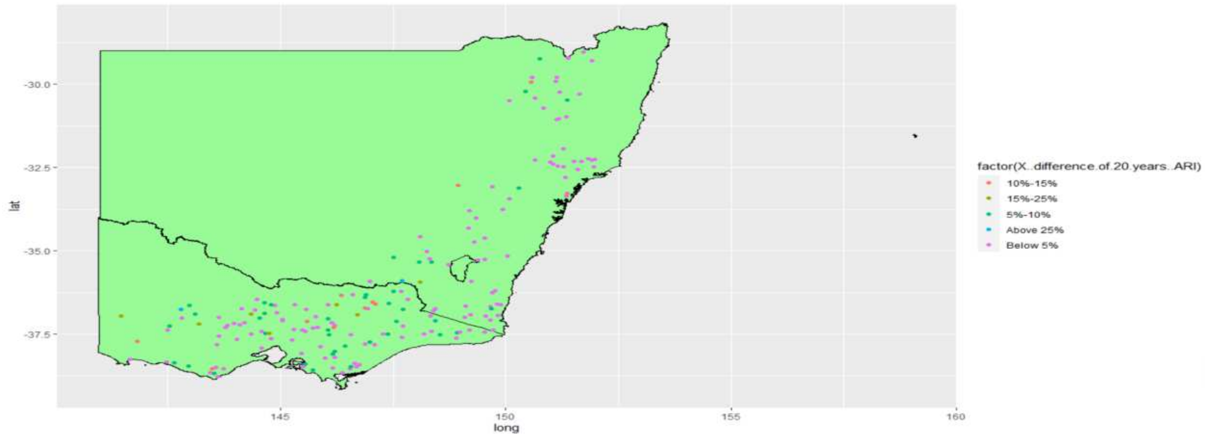


Figure 11 Geographical distribution of percentage differences between POT-ND and AM approaches (20 years ARI)

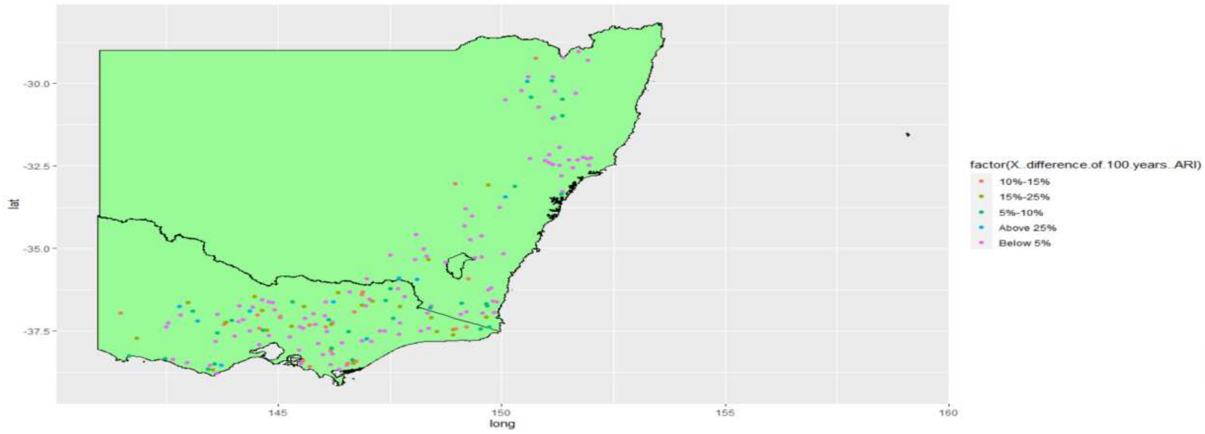
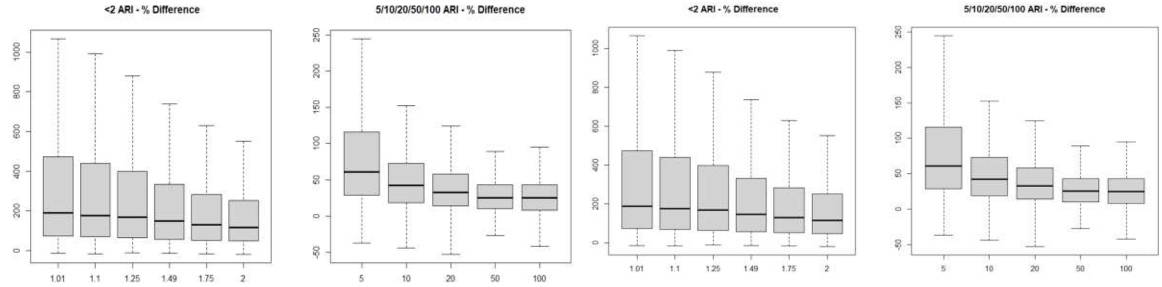


Figure 12 Geographical distribution of percentage differences between POT-ND and AM approaches (100 years ARI)



(a) *(b)*

Figure 13 (a) Comparison between the POT-ND parametric (b) POT-TS parametric and POT-non-parametric approaches– boxplots showing % difference

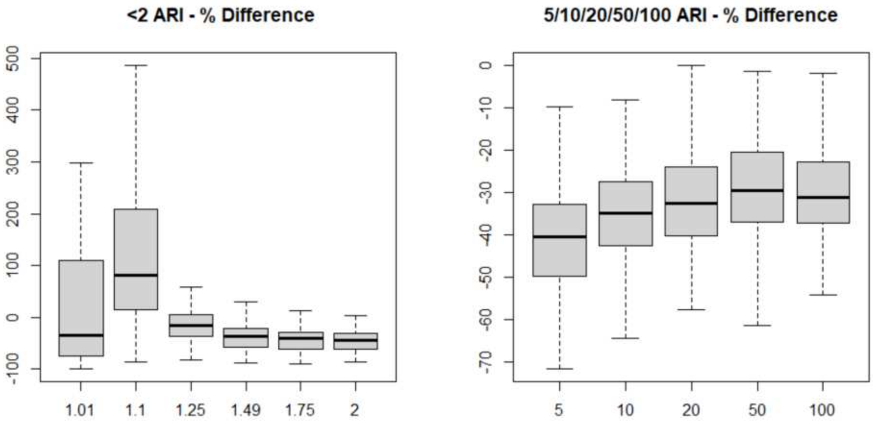


Figure 14 Comparison between POT-non-parametric and AM-GEV approaches – boxplot showing % difference

Table 1 Comparison between POT-ND and AM approaches

% Difference Based on Equation 6										
ARI	**POT-GP-10-MLE**	**POT-GP-10-PWMU**	**POT-GP-3-MLE**	**POT-GP-3-PWMU**	**POT-GP-5-MLE**	**POT-GP-5-PWMU**	**Mean**	**Median**	**SD**	**Preferred POT model**
1.01	-144%	-156%	-230%	-229%	-183%	-188%	-189%	-186%	0.36	POT-ND-GP-10-MLE
1.10	-509%	-585%	-1121%	-1100%	-807%	-832%	-826%	-820%	2.54	POT-ND-GP-10-MLE
1.25	86%	109%	297%	282%	198%	202%	196%	200%	0.87	POT-ND-GP-10-MLE
1.49	11%	20%	107%	96%	67%	66%	61%	67%	0.39	POT-ND-GP-10-MLE
1.75	-3%	2%	66%	57%	40%	38%	33%	39%	0.28	POT-ND-GP-10-PWMU
2.00	-8%	-4%	49%	41%	29%	28%	23%	28%	0.24	POT-ND-GP-10-PWMU
5.00	-15%	-12%	13%	10%	5%	6%	1%	5%	0.12	POT-ND-GP-5-MLE
10.00	-16%	-11%	2%	4%	-3%	0%	-4%	-2%	0.08	POT-ND-GP-5-PWMU
20.00	-17%	-11%	-6%	1%	-11%	-5%	-8%	-8%	0.06	POT-ND-GP-3-PWMU
50.00	-19%	-10%	-15%	-3%	-20%	-12%	-13%	-13%	0.06	POT-ND-GP-3-PWMU
100.00	-21%	-10%	-22%	-5%	-26%	-16%	-17%	-18%	0.08	POT-ND-GP-3-PWMU

Table 2(a) Overall site count for least % difference between POT-ND and AM approach

POT Model	Combined % difference (Count of sites)	Less than 2 years' ARI (Count of sites)	2 to 10 years' ARI (Count of sites)	20 & 50 years' ARI (Count of sites)	100 years' ARI (Count of sites)
POT3-ND	2	2	35	123	105
	1% of selected sites	1% of selected sites	19% of selected sites	65% of selected sites	56% of selected sites
POT5-ND	8	0	101	43	45
	4% of selected sites	0% of selected sites	54% of selected sites	23% of selected sites	24% of selected sites
POT10-ND	178	186	52	22	38
	95% of selected sites	99% of selected sites	28% of selected sites	12% of selected sites	20% of selected sites

Table 2(b) Overall site count for least % difference between POT-TS and AM approach

POT Model	Combined % difference (Count of sites)	Less than 2 years' ARI (Count of sites)	2 to 10 years' ARI (Count of sites)	20 & 50 years' ARI (Count of sites)	100 years' ARI (Count of sites)
POT3-TS	3	3	34	121	111
	2% of selected sites	2% of selected sites	18% of selected sites	64% of selected sites	59% of selected sites
POT5-TS	36	36	105	45	40
	19% of selected sites	19% of selected	56% of selected sites	24% of selected	21% of selected sites

		sites		sites	
POT10-TS	149 79% of selected sites	149 79% of selected sites	50 27% of selected sites	24 13% of selected sites	38 20% of selected sites

Table 3(a) Count of sites based on percentage difference between POT-ND and AM approaches

	1.01 ARI	**20 ARI**	**100 ARI**
Below 5% Difference	22	121	99
5%-10% Difference	15	54	44
10%-25% Difference	14	11	20
Above 25% Difference	137	2	25

Table 3(b) Count of sites based on percentage difference between POT-TS and AM approaches

	1.01 ARI	**20 ARI**	**100 ARI**
Below 5% Difference	22	121	104
5%-10% Difference	15	47	32
10%-25% Difference	14	11	20
Above 25% Difference	137	9	32

Table 4(a) Statistics of comparison between POT-ND and AM approaches

% Difference Based on Equation (6)				
ARI	**Mean**	**Median**	**SD**	**Preferred POT model**
1.01	660.08%	25.00%	3587.80%	POT-ND-GP-10
1.1	653.87%	259.00%	1529.25%	POT-ND-GP-10
1.25	84.22%	40.00%	127.91%	POT-ND-GP-10
1.49	20.17%	12.00%	24.00%	POT-ND-GP-10
1.75	12.18%	9.00%	13.16%	POT-ND-GP-10
2	9.90%	7.00%	10.23%	POT-ND-GP-5
5	5.44%	4.00%	6.77%	POT-ND-GP-5
10	5.02%	3.00%	7.45%	POT-ND-GP-5
20	5.56%	3.00%	9.66%	POT-ND-GP-5
50	7.91%	5.00%	13.65%	POT-ND-GP-5
100	9.71%	5.00%	17.18%	POT-ND-GP-3
Overall	145.00%	65.00%	345.00%	N/A

Table 4(b) Statistics of comparison between POT-TS and AM approaches

% Difference Based on Equation (6)				
ARI	**Mean**	**Median**	**SD**	**Preferred POT model**
1.01	660%	25%	3588%	POT-TS-GP-10
1.1	654%	259%	1529%	POT-TS-GP-10
1.25	84%	40%	128%	POT-TS-GP-10
1.49	20%	12%	24%	POT-TS-GP-10
1.75	12%	9%	13%	POT-TS-GP-10
2	10%	7%	10%	POT-TS-GP-10
5	5%	4%	7%	POT-TS-GP-5
10	5%	3%	7%	POT-TS-GP-5
20	6%	3%	10%	POT-TS-GP-3
50	8%	5%	14%	POT-TS-GP-3
100	10%	5%	17%	POT-TS-GP-3
Overall	134%	33%	486%	N/A

Table 5(a) Overall site count based on least % difference between the POT-ND parametric and POT-non-parametric approaches

POT Model	Combined % difference (Count of sites)	Less than 2 years' ARI (Count of sites)	2 to 10 years' ARI (Count of sites)	20 & 50 years' ARI (Count of sites)	100 years' ARI (Count of sites)
POT-ND-GP-3	2	1	2	4	28
	1%	1%	1%	2%	15%
POT-ND-GP-5	23	23	24	36	68
	12%	12%	13%	19%	36%
POT-ND-GP-10	163	164	162	148	92
	87%	87%	86%	79%	49%

Table 5(b) Overall site count based on least % difference between the POT-TS parametric and POT-non-parametric approaches

POT Model	Combined % difference (Count of sites)	Less than 2 years' ARI (Count of sites)	2 to 10 years' ARI (Count of sites)	20 & 50 years' ARI (Count of sites)	100 years' ARI (Count of sites)
POT-TS-GP-3	5	1	2	7	28
	3%	1%	1%	4%	15%
POT-TS-GP-5	19	0	8	29	55
	10%	0%	4%	15%	29%
POT-TS-GP-10	164	187	178	152	105
	87%	99%	95%	81%	56%

Table 6(a)Statistics of comparison between POT-ND parametric and non-parametric approaches

	Mean	SD	Median
Overall	106%	267%	46%
1-1.75 ARI	189%	559%	73%
2,5,10 ARI	52%	65%	29%
20,50 ARI	16%	13%	13%
100 ARI	13%	14%	8%

Table 6(b)Statistics of comparison between POT-TS parametric and non-parametric approaches

	Mean	SD	Median
Overall	104%	640%	21%
1-1.75 ARI	189%	940%	61%
2,5,10 ARI	51%	90%	20%
20,50 ARI	15%	16%	10%
100 ARI	13%	14%	8%

Table 7 Statistics of comparison between AM-GEV parametric and POT-non-parametric approaches

	Mean	SD	Median
Overall	46%	1098%	-33%
1-1.75 ARI	144%	1623%	-25%
2,5,10 ARI	-41%	16%	-39%
20,50 ARI	-30%	13%	-31%
100 ARI	-29%	16%	-31%

Figures

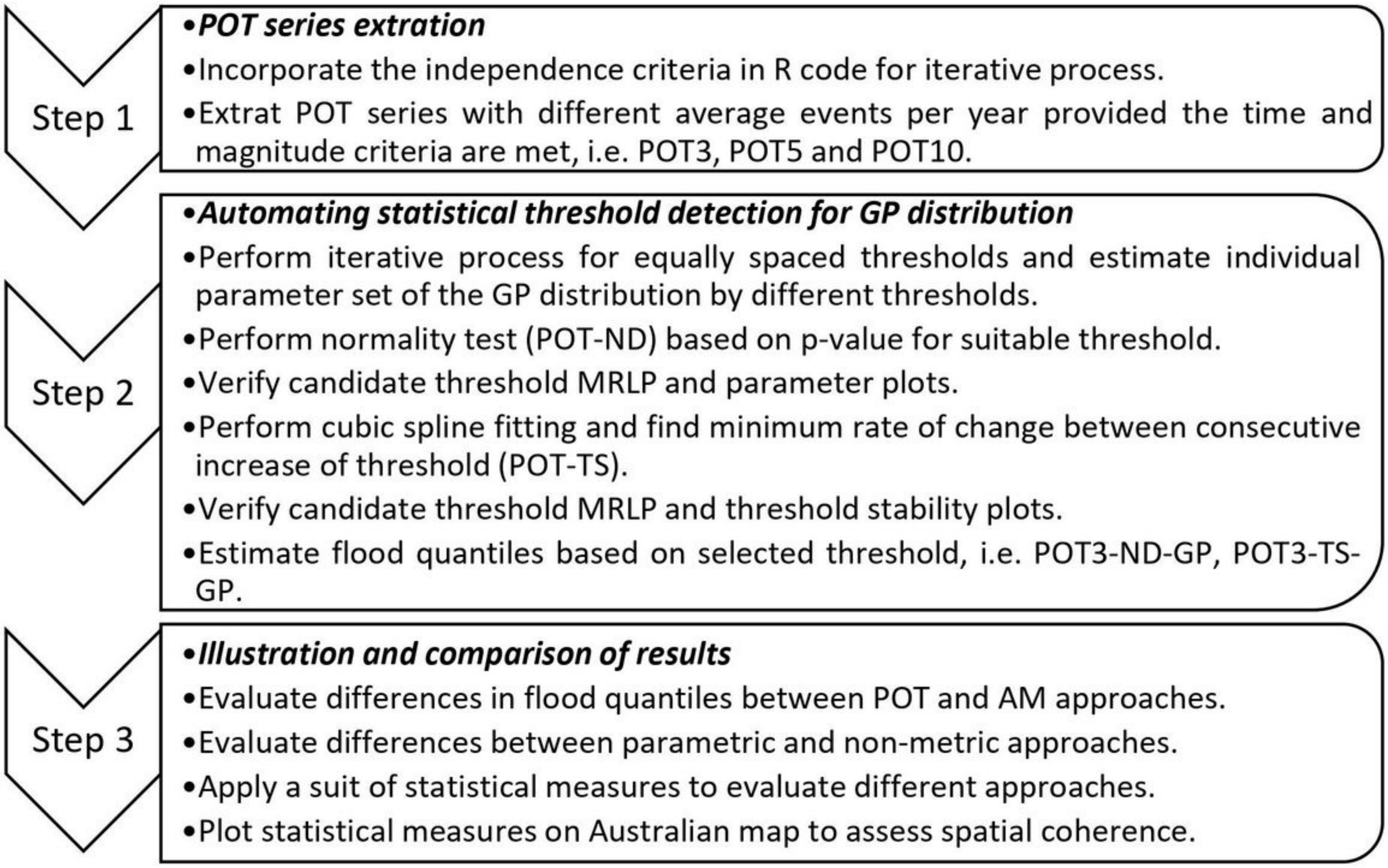


Figure 1

Illustration of adopted methodology

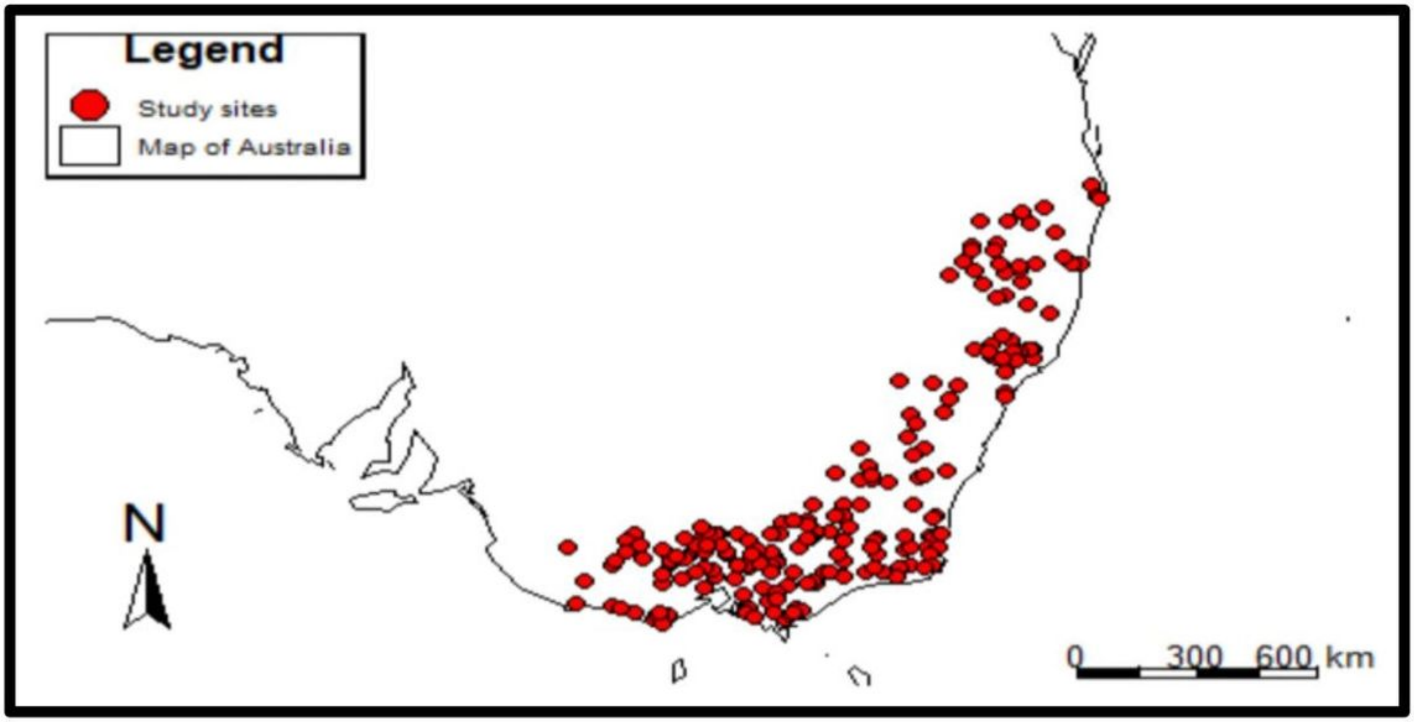


Figure 2

Distribution of selected stations Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

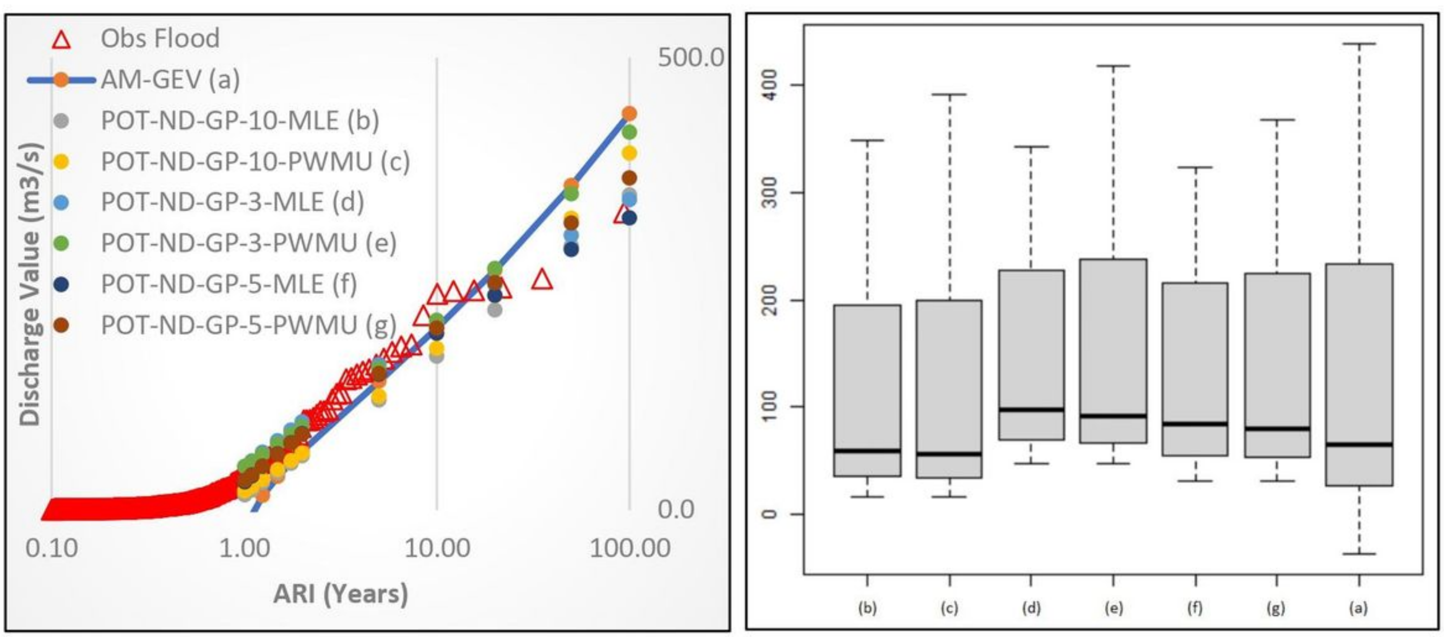


Figure 3

3(a) & 3(b) Comparison between POT-ND-GP (p-value=0.25) and AM-GEV approaches. Y-axis of the boxplots represents % difference estimated by equation 6.

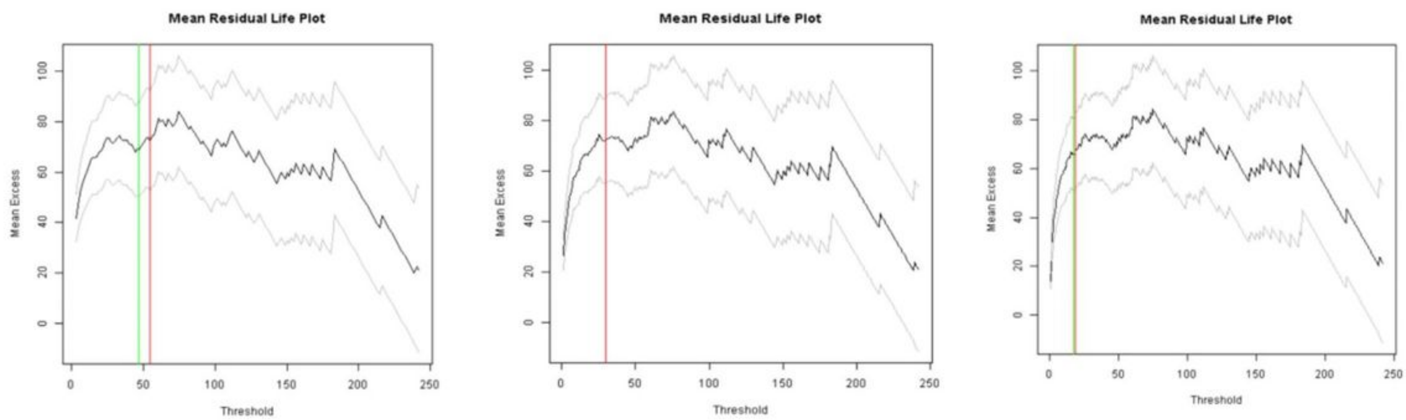


Figure 4

Mean residual life plot based on different p-values for POT-ND method

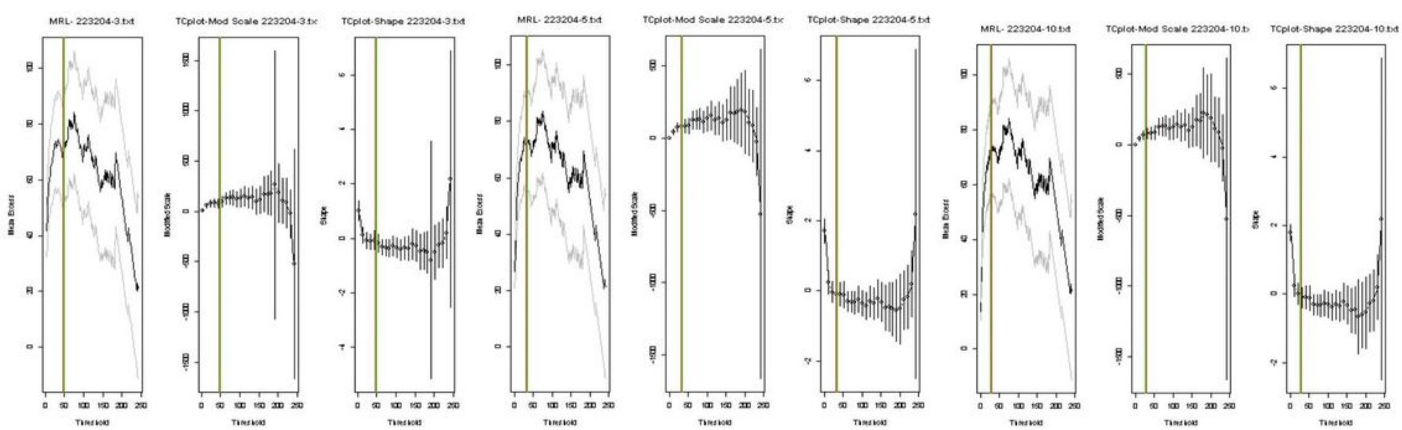


Figure 5

Estimated parameters plot with MRLP for POT-ND method

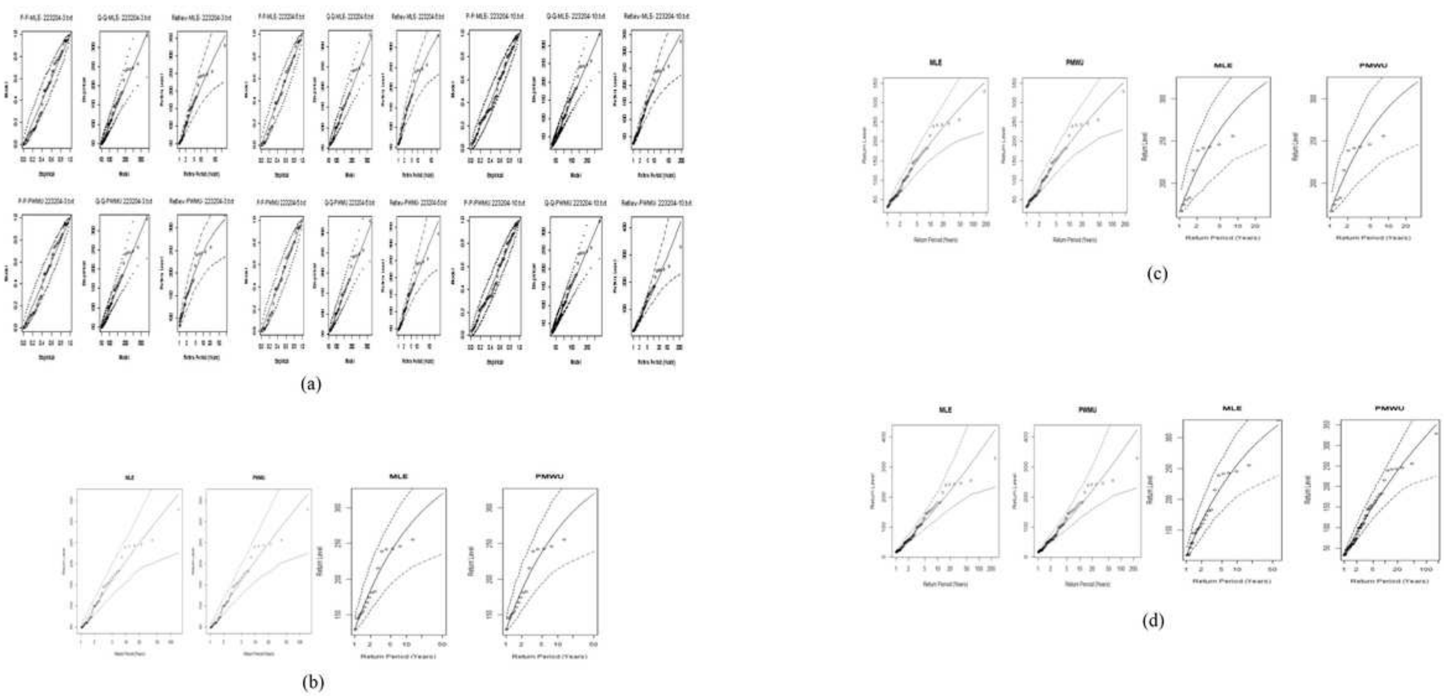


Figure 6

6(a) P-P, Q-Q and return level plots for different POT models using POT-ND method 6(b) Return level comparison between POT-ND-3 and POT-TS-3. From left to right, return level plots POT-ND-3-MLE, POT-ND-3-PWMU, POT-TS-3-MLE and POT-TS-3- PWMU 6(c) Return level comparison between POT-ND-3 and POT-TS-5. From left to right, return level plots POT-ND-5-MLE, POT-ND-5-PMWU, POT-TS-5-MLE and POT-TS-5- PWMU 6(d) Return level comparison between POT-ND-3 and POT-TS-10. From left to right, return level plots POT-ND-10-MLE, POT-ND-10-PWMU, POT-TS-10-MLE and POT-TS-10- PWMU

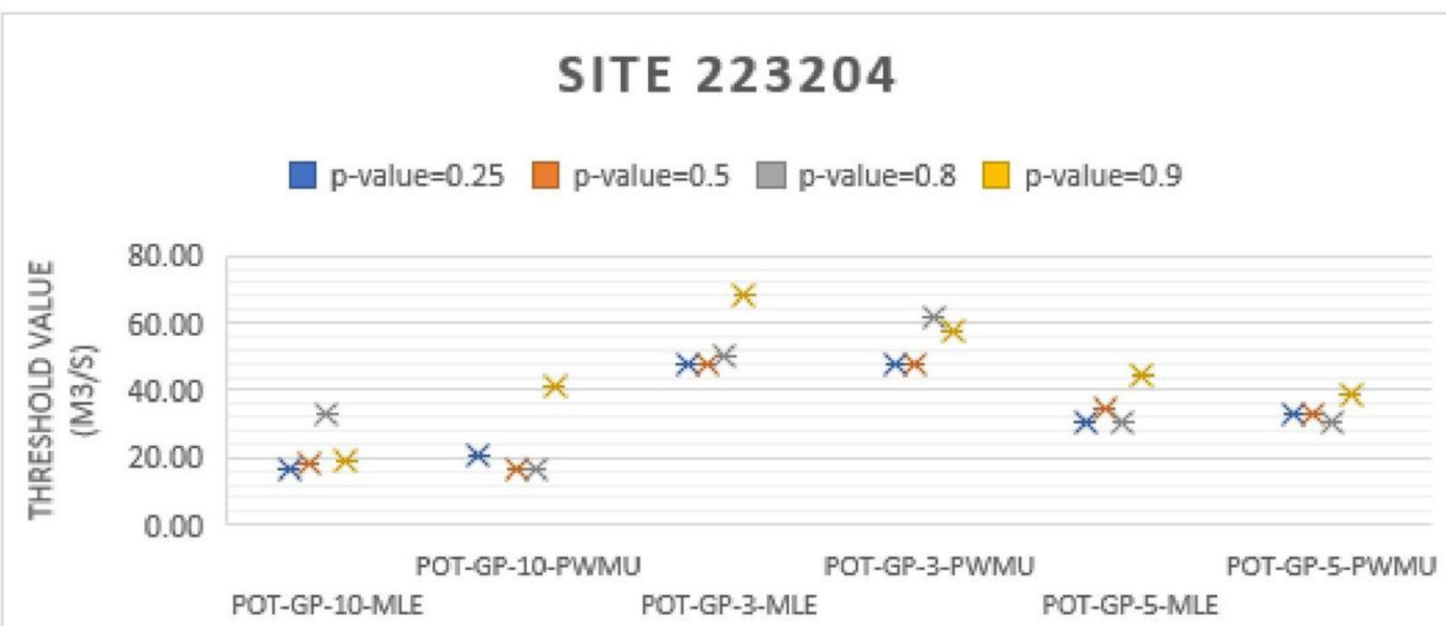


Figure 7

Threshold comparison based on varied p-value using POT-ND method

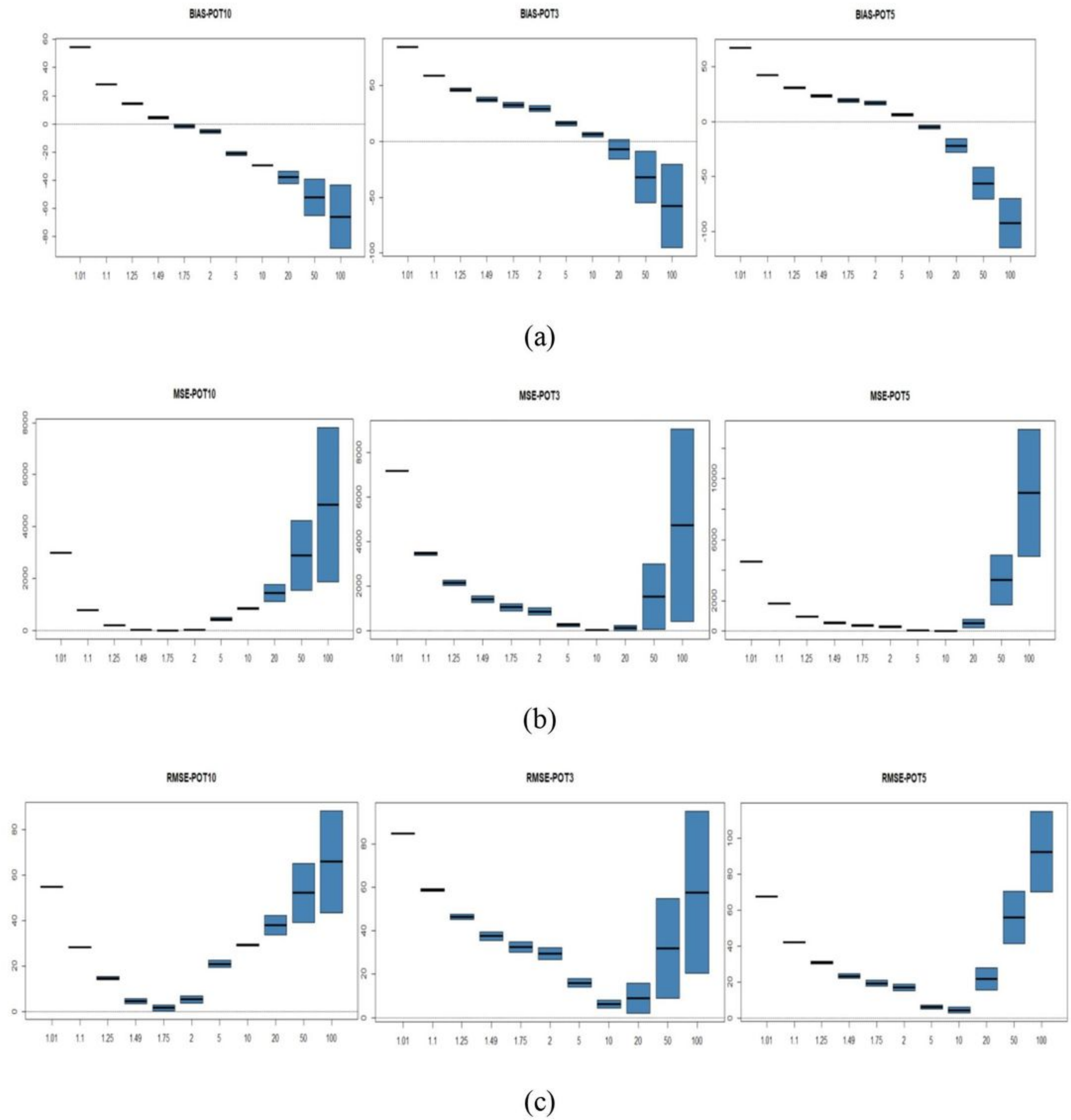


Figure 8

Verification Plot - (a) Bias Boxplot (b) MSE Boxplot (c) RMSE Box plot; from left to right, plots represent POT-ND-10, POT-ND-3 and POT-ND-5, respectively

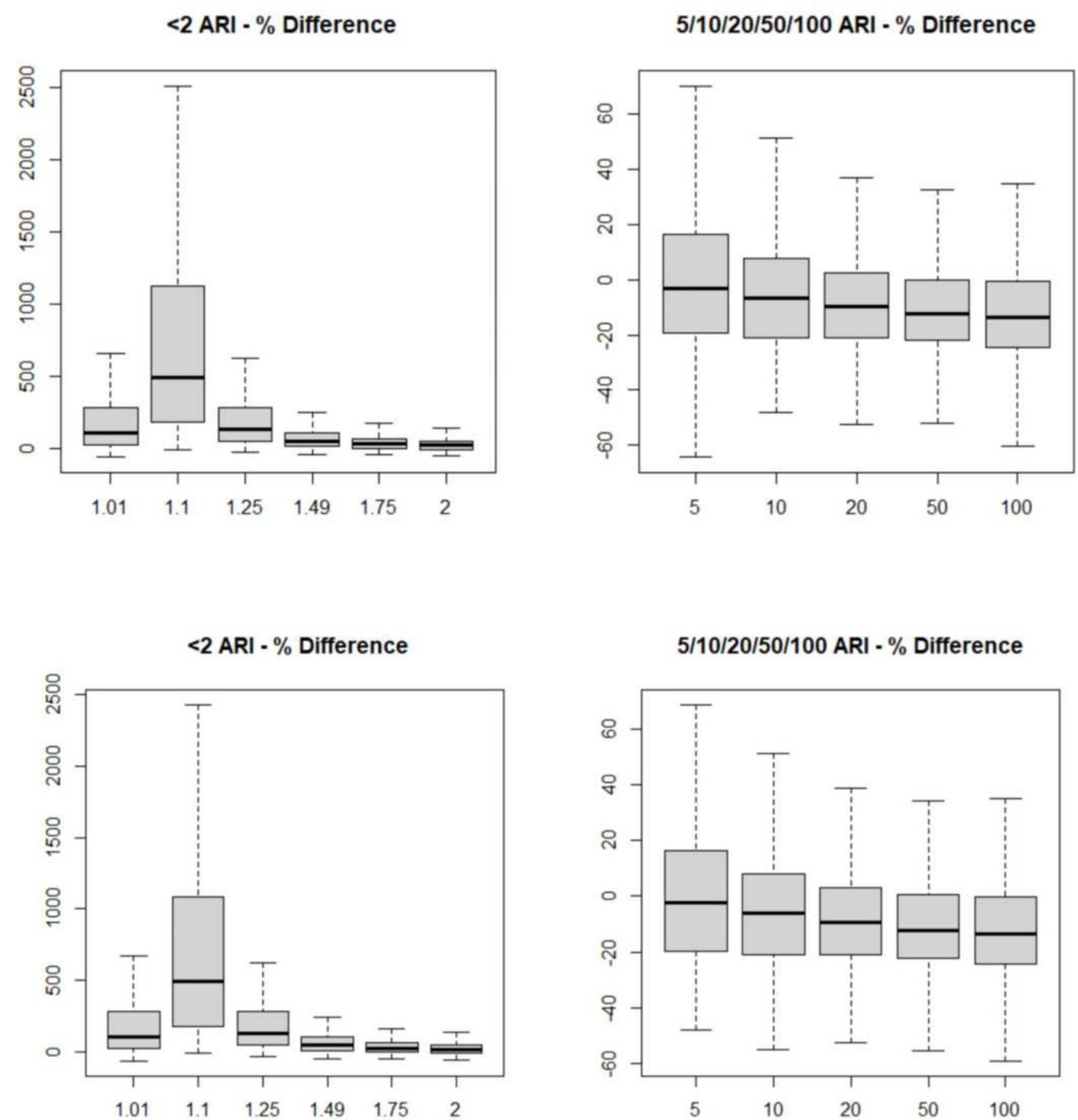


Figure 9

9(a) Comparison between POT-ND-GP and AM-GEV approaches – boxplots showing % difference 9(b) Comparison between POT-ND-GP and AM-GEV approaches – boxplots showing % difference

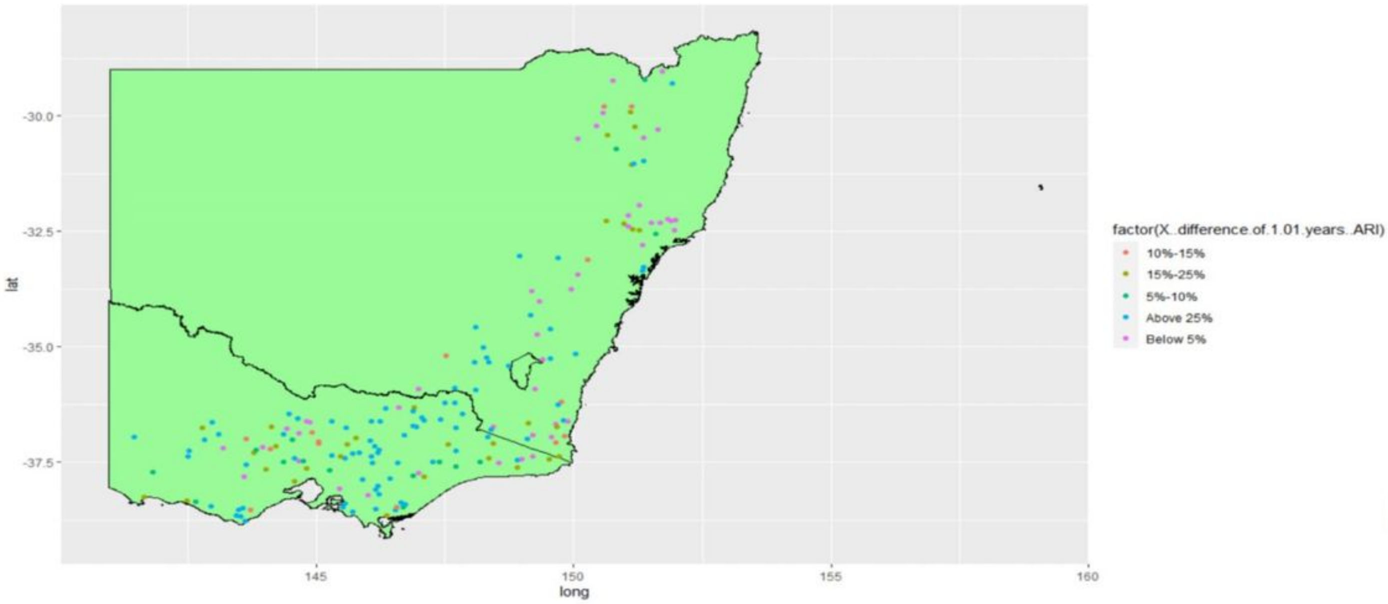


Figure 10

Geographical distribution of percentage differences between POT-ND and AM approaches (1.01 year ARI) Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

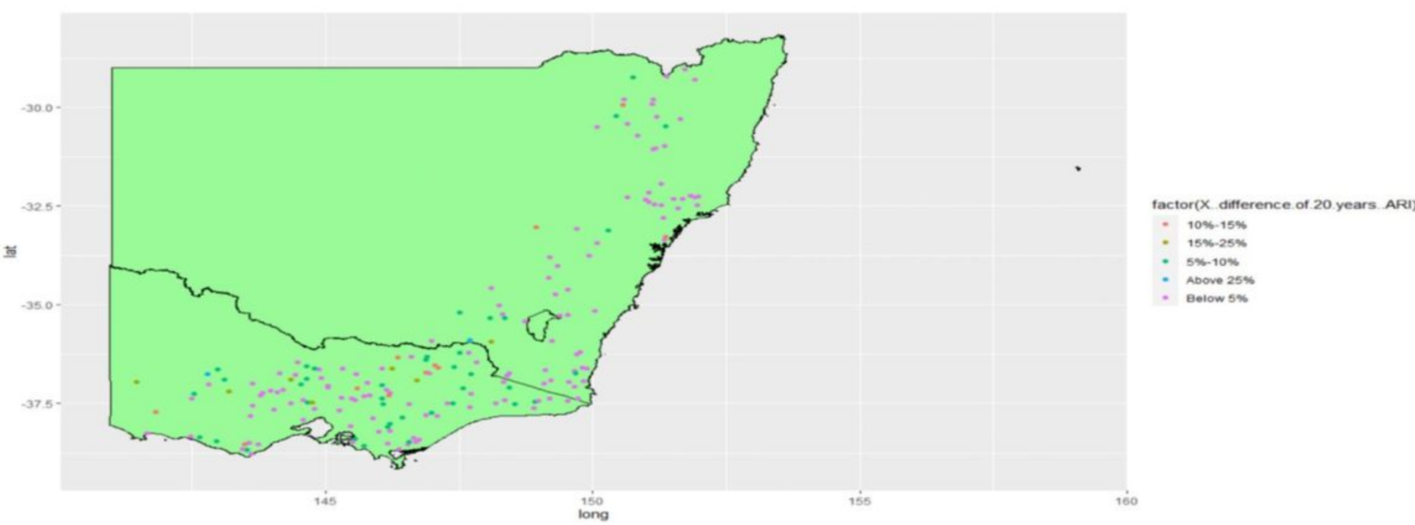


Figure 11

Geographical distribution of percentage differences between POT-ND and AM approaches (20 years ARI) Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

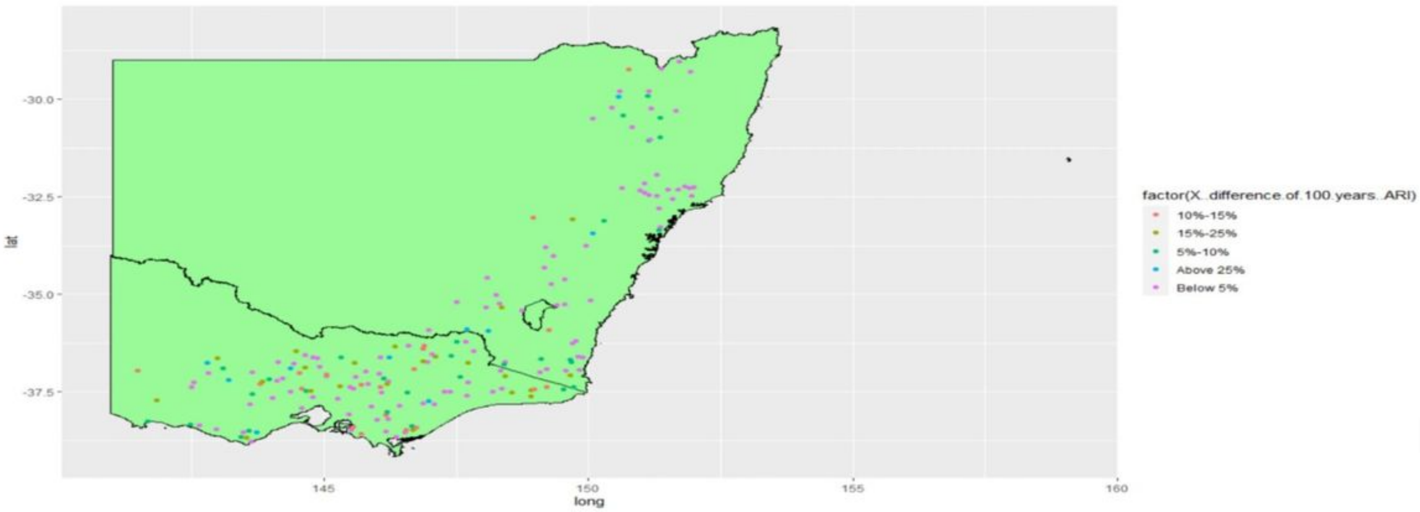


Figure 12

Geographical distribution of percentage differences between POT-ND and AM approaches (100 years ARI) Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

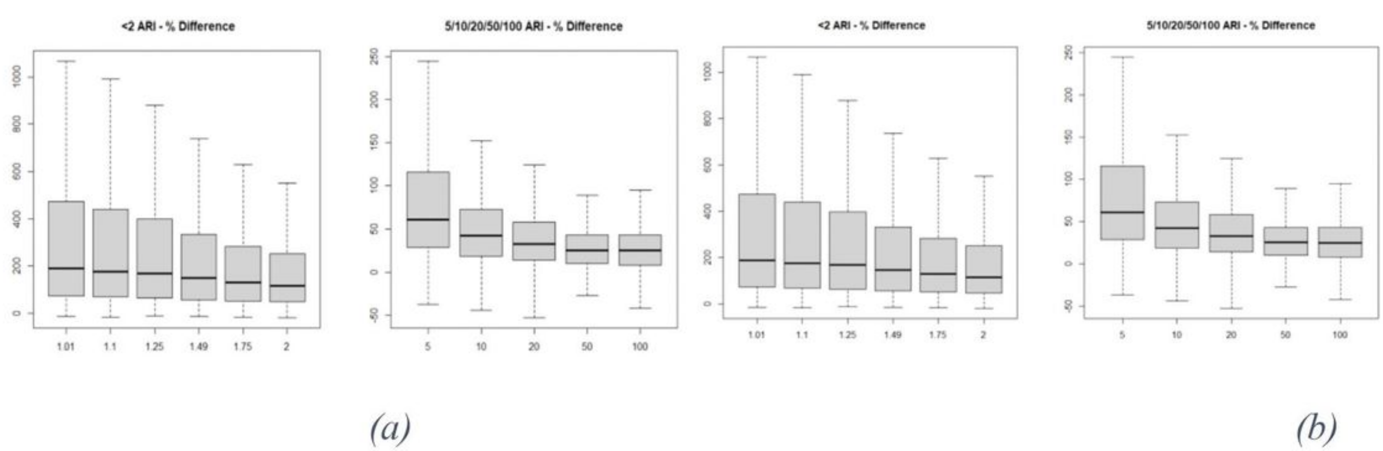


Figure 13

(a) Comparison between the POT-ND parametric (b) POT-TS parametric and POT-non-parametric approaches– boxplots showing % difference

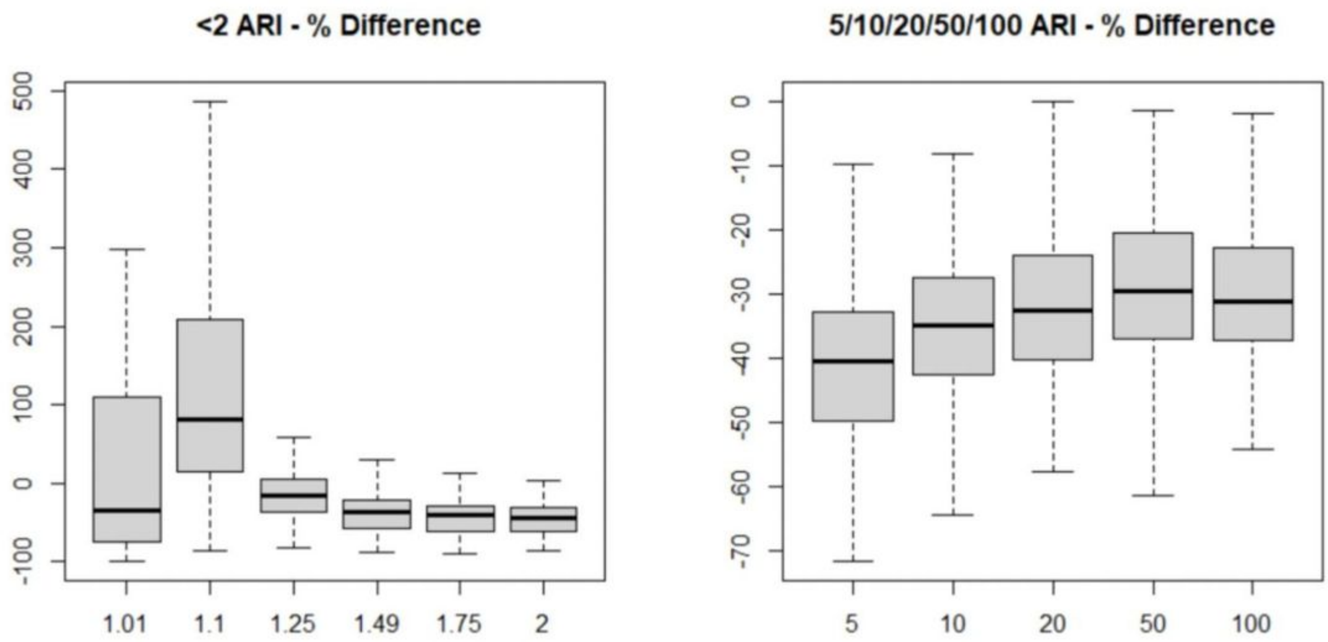


Figure 14

Comparison between POT-non-parametric and AM-GEV approaches – boxplot showing % difference